# EVOLUTIONARY GENOMICS OF CARNIVOROUS NEPENTHES PITCHER PLANTS

Mathias Scharmann

Diss. ETH No. 24464

DISS. ETH NO. 24464

# EVOLUTIONARY GENOMICS OF CARNIVOROUS *NEPENTHES* PITCHER PLANTS

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

MATHIAS SCHARMANN
Diplom-Biologe, Universität Würzburg

born 9 March 1986
citizen of Germany

accepted on the recommendation of

Prof. Dr. Alex Widmer, examiner
Prof. Dr. Christian Lexer, co-examiner
Prof. Dr. John Pannell, co-examiner

2017

Mathias Scharmann: *Evolutionary genomics of carnivorous Nepenthes pitcher plants*
© 2017
doi: 10.3929/ethz-b-000213059

Front cover: *N. ephippiata* upper pitcher, Hose Mountains, Sarawak, 2014
Back cover: upper pitcher of a hybrid involving *N. hemsleyana* and *N. rafflesiana* t.f., Tutong White Sands, Brunei, 2011
Photos M. Scharmann

Contents

Abstract vii
Zusammenfassung ix
General Introduction 1
Chapter 1: Divergence with gene flow among multiple sympatric carnivorous *Nepenthes* pitcher plants is linked to trap morphology 17
Abstract . . . . . 17
Introduction . . . . . 18
Materials and Methods . . . . . 21
Results . . . . . 26
Discussion . . . . . 32
References . . . . . 37
Chapter 1 Supporting Information S1, S2, S3, S4, S7 45
Chapter 2: Sex-determination and sex chromosomes are shared across the radiation of dioecious *Nepenthes* pitcher plants 57
Abstract . . . . . 57
Introduction . . . . . 57
Materials and Methods . . . . . 61
Results . . . . . 64
Discussion . . . . . 67
References . . . . . 72
Chapter 2 Supporting Information Methods S1, S2, S3, S4, S5 77
Chapter 3: Adaptive evolution of carnivory in *Nepenthes* pitcher plants: a comparative transcriptomics and proteomics perspective 87
Abstract . . . . . 87
Introduction . . . . . 88
Results . . . . . 90
Discussion . . . . . 98
Methods . . . . . 100
References . . . . . 108
General Discussion 117
Acknowledgments 125
Curriculum Vitae 127

# Abstract

*Nepenthes* is a species-rich genus of dioecious, carnivorous plants, that has evolved a fascinating diversity of pitcher traps that vary in size, shape and structure and attract different types of prey. At many sites in the natural distribution range of the genus, multiple species can be found together. Although hybrids can be formed experimentally and sometimes arise in nature, species apparently coexist in sympatry and display species-specific traits primarily in their pitcher traps. This may suggest that the trap differences represent adaptations that are maintained by natural selection in the face of gene flow. To address this question, I surveyed population genomic patterns in seven *Nepenthes* species with disparate traits that co-occur at two sites. I tested the hypothesis that interspecific gene flow occurs in these communities, which may indicate a role for divergent selection in the maintenance of separate species. On the basis of thousands of single-nucleotide polymorphisms (SNPs) sampled throughout the genomes, I found that these sympatric species are genetically distinct but that reproductive isolation is not complete under natural conditions. A phylogenomic approach revealed that these species populations share genetic variation to an extent that cannot be explained by retained ancestral variation alone. Coalescent modelling suggested that gene flow is not (only) ancient but more probably ongoing between multiple species pairs. The selective agents enforcing partial genomic isolation against gene flow appear to be linked to the pitcher traps, as the morphological differentiation of this organ in a young species pair was higher than expected from neutral genetic differentiation. I conclude this chapter by presenting the hypothesis that speciation in *Nepenthes* is promoted by physiological cost-benefit trade-offs in the carnivorous syndrome.

The mechanism of sex determination in *Nepenthes* is currently unknown. The analysis of chromosome morphology does not provide insights because all chromosomes are similar in the hitherto examined *Nepenthes* species. Further, like many other perennial plants, *Nepenthes* is not amenable to standard crossing and segregation analysis because of a long juvenile phase and challenging cultivation. An alternative approach to reveal genetic sex determination is to scan natural populations for molecular markers that are either associated with sex or are specific to one sex. I developed a statistical method to distinguish between stochastic noise and true sex-specificity in population genomic data, which overcomes existing problems that have so far hindered a wider application of this approach. I found shared male-specific DNA sequence markers but an absence of female-specific markers in three distantly related *Nepenthes* taxa, suggesting the existence of Y-chromosomes and a common origin of dioecy. Sex-linked markers matched several expressed transcripts from a male inflorescence, among them candidate

genes for pollen development and floral organ identity. This relatively old, yet homomorphic XY-system with a Y-specific region that ranges among the smallest known in plants contributes to our understanding of the diversity of plant sex chromosomes.

Based on my findings that *Nepenthes* pitcher traps may be under divergent selection, I focussed on the evolution of genes underlying prey digestion. Through the combination of proteomics with transcriptomics for 12 species, I identified secreted digestive enzymes and other proteins in pitcher fluids, unveiling unexpectedly high diversity within and between species. Analysis of differential gene expression between fed pitchers, control pitchers, and non-carnivorous leaves point towards a major physiological shift induced by feeding. During the *Nepenthes* radiation, thousands of genes experienced positive selection on amino acid substitutions and shifted in their expression. Among both of these groups, secreted pitcher fluid proteins were over-represented, in contrast to other genes that are upregulated upon feeding. However, these feeding-related genes were enriched among the differentially expressed genes in a young species pair. My results show that positive selection at the structural gene level, together with regulatory divergence occurred extensively during the *Nepenthes* radiation, and are consistent with the hypothesis of adaptive evolution of carnivory.

# Zusammenfassung

*Nepenthes* ist eine artenreiche Gattung diözischer, fleischfressender Pflanzen, die eine faszinierende Vielfalt von Kannenfallen entwickelt hat, welche sich in Grösse, Form und Struktur unterscheiden und verschiedene Beutetypen anlocken. An vielen Standorten innerhalb des natürlichen Verbreitungsgebietes der Gattung kommen mehrere Arten zusammen vor. Obwohl Hybriden künstlich erzeugt werden können und manchmal in der Natur entstehen, bestehen die Arten offensichtlich in Sympatrie nebeneinander und zeigen artspezifische Eigenschaften vor Allem in ihren Kannenfallen. Dies könnte darauf hinweisen, dass die Unterschiede in den Fallen Anpassungen darstellen, die durch natürliche Selektion trotz Genfluss aufrechterhalten werden. Um diese Frage zu beantworten, erfasste ich populationsgenomische Muster in sieben *Nepenthes* Arten mit ungleichen Eigenschaften, die an zwei Standorten zusammen vorkommen. Ich prüfte die Hypothese, dass zwischenartlicher Genfluss in diesen Artgemeinschaften auftritt. Dies kann ein Hinweis darauf sein, dass divergente Selektion in der Aufrechterhaltung getrennter Arten eine Rolle spielt. Anhand tausender über das Genom verteilter Einzelnukleotid-Polymorphismen (SNPs) fand ich heraus, dass diese sympatrischen Arten zwar genetisch verschieden, jedoch unter natürlichen Bedingungen nicht vollständig durch Fortpflanzungsbarrieren getrennt sind. Ein phylogenomischer Ansatz zeigte, dass die zwischen Artenpopulationen geteilte genetische Variation ein Ausmass hat, das nicht allein durch Beibehalten von genetischer Variation des gemeinsamen Vorfahren erklärbar ist. Koaleszenzmodellierung liess den Schluss zu, dass dieser Genfluss nicht (nur) alt ist, sondern mit höherer Wahrscheinlichkeit zwischen mehreren Artenpaaren gegenwärtig andauert. Die Selektionskräfte, die eine teilweise genomische Trennung trotz Genfluss erzwingen, scheinen mit den Kannenfallen in Verbindung zu stehen, weil die morphologischen Unterschiede dieser Organe in einem jungen Artenpaar höher ausfielen als durch neutrale genetische Unterschiede zu erwarten. Ich beschliesse dieses Kapitel mit der Darstellung der Hypothese, dass die Artbildung in *Nepenthes* durch gegenläufige Abhängigkeiten zwischen physiologischen Kosten und Nutzen der Karnivorie vorangetrieben wird.

Der Geschlechtsbestimmungsmechanismus von *Nepenthes* ist aktuell unbekannt. Die Untersuchung von Chromosomenmorphologie ergibt keine Anhaltspunkte, weil bei den bisher untersuchten *Nepenthes* Arten alle Chromosomen ähnlich sind. Zudem ist *Nepenthes* wegen des langen Juvenilstadiums und hoher Kulturansprüche nicht einfach zugänglich für die Standardmethoden Kreuzung und Vererbungsanalyse, wie auch viele andere mehrjährige Pflanzen. Alternativ dazu kann eine genetische Grundlage der Geschlechtsbestimmung gefunden werden indem man in natürlichen Populationen nach molekularen Markern sucht, die

Allelfrequenzunterschiede zwischen den Geschlechtern zeigen oder in nur einem der Geschlechter vorkommen (Geschlechts-Spezifität). Ich entwickelte eine statistische Methode um in populationsgenomischen Daten zwischen Zufallsrauschen und tatsächlicher Geschlechts-Spezifität zu unterscheiden, was bestehende Probleme löst die eine verbreitete Anwendung dieses Ansatzes bisher verhindert haben. Ich fand gemeinsame, männchen-spezifische DNA Sequenzmarker jedoch keine weibchen-spezifischen Marker in drei entfernt verwandten *Nepenthes* Arten. Dies legt die Existenz von Y-Chromosomen und einen gemeinsamen Ursprung der Diözie nahe. Geschlechtsabhängige Marker passten zu einigen exprimierten Genen eines männlichen Blütenstandes, darunter Kandidatengene für die Pollenentwicklung und die Identitat von Blütenorganen. Dieses relativ alte aber homomorphe XY-System, mit einer Y-spezifischen Region die unter den kleinsten der von Pflanzen bekannten ist, trägt zum Verständnis der Vielfalt pflanzlicher Geschlechtschromosomen bei.

Ausgehend von meinen Erkenntnissen, dass die Kannenfallen von *Nepenthes* unter divergenter Selektion sein könnten, betrachtete ich die Evolution von Genen, die dem Beuteverdau zugrunde liegen. Ich bestimmte mithilfe der Kombination von Proteomik und Transkriptomik die sekretorischen Verdauungsenzyme und weitere Proteine in Kannenflüssigkeiten von 12 Arten, wobei eine unerwartet hohe Vielfalt innerhalb und zwischen Arten enthüllt wurde. Die Untersuchung von Genexpressionsunterschieden zwischen gefütterten Kannen, Kontroll-Kannen sowie nicht-fleischfressenden Blättern weist darauf hin, dass während des Fressens bedeutende physiologische Veränderungen ausgelöst werden. Innerhalb der Radiation der Gattung *Nepenthes* waren tausende von Genen unter positiver Selektion für Aminosäuresubstitutionen und veränderten ihre Expression. In diesen beiden Kategorien sind sekretorische Proteine der Kannenflüssigkeit überproportional vertreten, im Gegensatz zu anderen Genen die während des Fressens heraufreguliert werden. Jedoch sind diese am Fressen beteiligten Gene angereichert unter denjenigen Genen, welche ebenfalls in einem jungen Artenpaar unterschiedlich exprimiert werden. Meine Ergebnisse zeigen, dass positive Selektion auf der Ebene der Genstruktur sowie auf der regulatorischen Ebene innerhalb der *Nepenthes*-Radiation verbreitet auftrat, und stimmen mit der Hypothese von adaptiver Evolution der Karnivorie überein.

# General Introduction

The evolution of organismic diversity is the result of a complex interplay between stochastic and selective forces. Understanding how and why species diverge, adapt, and interact during this process requires integrative consideration of both aspects. Random mutation and genetic drift are sufficient to explain most of the variation in DNA and protein sequences (Kimura 1983), the most basic level of biodiversity. However, the relative importance of selection rises with increasing biological complexity from DNA over proteins to organisms. Between species, most phenotypic diversity may be consistent with selection and adaptations (Rieseberg et al. 2002), as envisaged by Darwin (1859). However, the subtle differences between closely related but not identical organisms, such as males and females of a single species, or incipient sister species, can be challenging to explain. This thesis takes a molecular genomic perspective to study organismic diversity in two particularly enigmatic but important processes, the evolution of species and of separate sexes.

## Speciation

Novel evolutionary lineages arise through speciation, a process during which barriers to genetic exchange other than spatial separation (reproductive isolation, RI) evolve (Baack et al. 2015). Speciation can be framed as a population genetic process that starts from a single sexual population in panmixia, and may complete with two differentiated populations that show only restricted or no genetic exchange between them, then termed species. This fundamental biological process is at the heart of evolutionary research (Stebbins 1950; Grant 1981; Coyne & Orr 2004; Rieseberg & Willis 2007), and many open questions remain (e.g. The Marie Curie SPECIATION Network 2012). Recent views put the realised level of gene flow in the centre of the discussion, which allows some generalisations, although the details of RI differ from case to case and are infinitely diverse. However, speciation and the evolution of RI can be summarised in terms of three continuous axes: spatial differentiation, ecological differentiation (e.g. contrasting environments), and mating differentiation (Dieckmann et al. 2004). Differentiation can advance along multiple routes within this framework, driven by dispersal, genetic drift or selection. Certainly, both neutrality and selection can contribute to speciation. Theoretical models of speciation have dramatically improved the understanding of the conditions under which speciation can be expected, and have made predictions about associated biological patterns from the scale of genes to ecological communities (Gavrilets 2014). However, it is still unknown how the most relevant factors – selection, genetic drift and dispersal – interact in

real organisms and whether any of them dominate in nature, especially during evolutionary radiations that produce multiple, ecologically co-existing species.

The early stages of speciation, when RI is still incomplete, are most suitable to gain deeper empirical insight into the process. When the geographic setting allows for contact, hybridisation can naturally occur between incipient species, and repeated back-crossing can lead to introgression of certain genomic regions and phenotypes (e.g. Wu 2001; Harrison & Larson 2014). Hybrid zones or genetic contact zones are natural experiments that can be used to elucidate the roles of selection, drift and genome architecture in speciation, and allow analysis of particular traits and their underlying genes (Abbott et al. 2013; Cruickshank & Hahn 2014). One central expectation in such situations is that the homogenising effect of gene flow (introgression) should erode neutral differentiation between species and expose targets of diversifying selection. Genome-wide assessments of diversity and divergence in populations undergoing speciation ('speciation genomics'), in combination with phenotypic studies and breeding experiments, allow insights into the evolution of RI in ever greater detail (Seehausen et al. 2014). These data help to test manifold hypothesis, such as the heterogeneity of divergence along genomes, the distribution of gene flow, the genes underlying RI and associated with divergent selection, and in which temporal sequence the three reproductive barrier classes prezygotic, postzygotic-intrinsic and postzygotic-extrinsic (mediated by the environment or biological interactions) build up during speciation.

## Molecular evolution and evolutionary radiations

Under suitable conditions, some evolutionary lineages undergo multiple, possibly rapid speciation events and generate many genetically and phenotypically differentiated species (Schluter 2000; Seehausen 2004). While species diversity per se can be readily discernible, the inference of adaptive evolution, and the identification of organismic traits affected by directional selection during radiations are challenging and require consideration of many aspects from phylogeny to ecology (Schluter 2000; Losos & Mahler 2010). It has been argued that some evolutionary radiations could be described as non-adaptive, because genetic drift facilitated by geographic isolation, or sexual selection, rather than adaptation for disparate ecological and physiological phenotypes can be an alternative explanation for rapid speciation (e.g. Givnish 2015).

Comparative phylogenomics or transcriptomics, in particular tests for molecular signatures of adaptations (Nielsen 2005), provide an approach that complements the conventional methods to test for adaptations in radiations (Schluter 2000; Losos & Mahler 2010). One of the ideas explored in this thesis consists of a transcriptome-wide scan for molecular signatures of adaptation during a species radiation, in combination with knowledge about the genes underlying interesting organismic traits that possibly underwent adaptive evolution. It is hypothesised that this strategy, which was applied by only few previous studies (Roux et al.

2014; Brawand et al. 2014; Pease et al. 2016), may directly indicate whether adaptations occurred in evolutionary radiations, and could reveal whether certain traits disproportionally contributed to them.

Protein coding genes can retain signatures of past selection regimes (e.g. Yang 2014). However, this potential is rarely leveraged to understand evolutionary radiations of organisms. First, most studies of coding sequence evolution focus on single genes or gene families, rather than taking a genome-wide perspective. The latter could allow an unbiased discovery of gene functions and pathways that preferentially experienced certain selection regimes. Second, tests of selection on coding sequences are predominantly conducted over deep phylogenetic scales (e.g. plants, mammals, etc.). However, molecular signatures of selection among deeply diverged organisms probably evolved long after speciation was completed, and thus contribute little to an understanding of the processes during the possibly rapid evolutionary radiations that generated most biodiversity. A pioneering study by Kapralov et al. (2013) found that diversifying selection on protein sequences played a significant role in an island adaptive radiation.

In addition to coding sequence evolution, regulatory changes at the level of gene expression are now recognised as important, perhaps dominating mechanisms underlying phenotypic divergence (Romero et al. 2012). A pre-genomics era study by Barrier et al. (2001) indicated a prominent role of regulatory genes in an adaptive radiation, but the concept has only very recently been explored over a large set of genes and closely related species (Nevado et al. 2016).

## Dioecy and plant sex determination

The perhaps most striking polymorphism that can be observed within populations of some but not all sexual species is the difference between sexes (genders). Reproduction with outbreeding (sexuality) is a characteristic feature of most eukaryotic organisms, because segregation and recombination of both detrimental and advantageous genetic variants provide superior long-term fitness according to theoretical and empirical results (Otto 2009). However, sexual reproduction sets the stage for parental conflict over investment into offspring, following a quantity-quality trade-off (sexual conflict). In most systems, two alternative strategies (genders or sexes) can stably coexist because their fitness is density-dependent: one with cheap but numerous gametes (male, producing sperm cells) and another with expensive but few gametes (female, producing ovules; anisogamy, Parker et al. 1972). A further consequence of this phenomenon is sexual selection, typically in the form of sperm cells competing for a limited number of ovules, while females, or the female function in hermaphrodites, can potentially choose high quality mates from the oversupply.

There is remarkable variation in the degree to which individuals express either one or both of these functions (sex allocation, Charnov 1982). Most plants pursue both male and female strategies (hermaphroditism), but c. 75% of all Angiosperm families also contain dioecious species (Renner & Ricklefs 1995). Which factors

promote the maintenance of hermaphroditism and dioecy in plants respectively transitions between these mating systems, and their relative importance, is a an active research field with a long history (Darwin 1877; Charlesworth & Charlesworth 1978; Charlesworth 1999; Käfer et al. 2017).

Inseparably connected to sex allocation are sex determination mechanisms, which are highly diverse (Bachtrog et al. 2014) and can act at the level of organs or entire organisms. Sex may be heritable (genetic determination), controlled by the environment, or both (Rodrigues et al. 2015). Only a small fraction of dioecious plants have been investigated for their sex determination system, and a causal gene is known in only one plant species (persimmon, Akagi et al. 2014). One conspicuous genetic mechanism of sex determination involves sex chromosomes. These are genomic regions with reduced recombination, and may or may not be detectable in karyotypes as unpaired or absent chromosomes and chromosome parts. Their divergence is thought to be initiated by the loss of recombination in a pair of (ancestral) autosomes, promoted by selection for the joint inheritance of sex determining loci and sexually antagonistic alleles (Charlesworth 1991), or by meiotic drivers (Úbeda et al. 2015). Notably, very few additional plant sex chromosomes were identified since the influential work of Westergaard (1958). This limits comparative studies and the understanding of fundamental aspects of dioecy and genome evolution, particularly the factors leading to homo- or heteromorphic sex chromosomes. More independent examples are required to better understand how dioecy evolves, how it is genetically controlled, and which general patterns exist in plant sex chromosomes and sex determination systems (Ming et al. 2011; Moore et al. 2016).

Dioecy and sex chromosomes can potentially affect population divergence and speciation in several ways. First, sex chromosomes recombine less during meiosis than autosomes, and exhibit at least partially reduced effective population sizes. For these and other reasons, they accrue genetic differentiation not only between sexes but as a side-effect also between populations or incipient species, at a faster rate than autosomes. Sex-chromosomes may thus enhance the evolution of reproductive isolation (Haldane's rule and the large X-effect, Presgraves 2008).

Second, sexual selection in the form of male-male competition for females is well characterised as a driver of the evolution of mating-success, and thought to indirectly enhance the evolution of prezygotic RI and thus speciation. Sexual selection in plants mostly occurs when pollen tubes, after germinating on the stigma, race to fertilise a limited number of ovules (Moore & Pannell 2011). Some evidence has been collected to support the hypothesis that the strength of sexual selection increases the level of interspecific reproductive isolation in plants (L. Rieseberg, reported in Lafon-Placette et al. 2016). Hence, one could hypothesise that dioecy, because it could be a consequence of sexual selection in hermaphrodites (Willson 1979), correlates with the evolution of pre-zygotic barriers promoted by sexual selection.

## *Nepenthes* pitcher plants

Plants of the genus *Nepenthes* (Nepenthaceae, Caryophyllales) are vines or shrubs that produce strongly modified leaves with a proximal part that appears like a conventional flat leaf, and a peculiar distal part that appears rolled up into an upright positioned tube with an opening at the top - the pitcher (Juniper et al. 1989; Clarke 1997, 2001; Cheek & Jebb 2001; McPherson 2009, 2011). All leaves except the cotyledons develop pitchers. These unique organs are traps that serve to sequester nutrients from the plant's environment, usually in the form of insects which are attracted, then drowned in the fluid-filled cavity, and digested for the release and absorption of nitrogen, phosphorus and other elements (Juniper et al. 1989; Moran & Clarke 2010). *Nepenthes* are photo-autotrophic and have functional roots (Pavlovič et al. 2010), like other carnivorous plants. Minerals extracted from prey are used to enhance photosynthesis (Pavlovič et al. 2009, 2011), which gives carnivorous plants an advantage when nutrients in the soil are scarce. For *Nepenthes*, carnivory may be essential under natural conditions, as they cease growth and do not flower when experimentally deprived of prey capture (Moran & Moran 1998). This reveals that carnivory can be adaptive, but presumably comes with a high energetic cost and constitutes a viable eco-physiological strategy only under a narrow set of conditions (e.g. low soil nutrient availability, low competition for light, Givnish et al. 1984; Ellison 2006; Pavlovič & Saganová 2015).

While they do not actively move, *Nepenthes* pitchers are highly derived, complex organs with a plethora of physiological and morphological adaptations involved in the process of carnivory. Prey is lured to the traps by extrafloral nectar, flower- or fruit-like odour, and colouration (Moran & Clarke 2010), and several biomechanical devices cause animals to slip and fall into the cavity, and prevent their escape (Gaume et al. 2002; Bohn & Federle 2004; Gorb et al. 2004, 2005; Gaume & Forterre 2007; Bauer et al. 2008, 2015; Bauer & Federle 2009; Scholz et al. 2010; Moran et al. 2012). Once immersed in the acidic fluid at the base of the traps, animals quickly drown and digestion commences. Triggered by substances released from prey, the plant actively participates in digestion by lowering the fluid pH and secretion of a cocktail of hydrolytic enzymes (Athauda et al. 1998, 2004; Eilenberg et al. 2006; Hatano & Hamada 2008, 2012; Rottloff et al. 2016; Lee et al. 2016; Fukushima et al. 2017). The molecular machinery of plant carnivory appears to be derived from defence mechanisms against pathogens and herbivores that also exist in non-carnivorous plants (Renner & Specht 2012; Bemm et al. 2016; Fukushima et al. 2017). Moreover, prey digestion in *Nepenthes* involves a diverse saprotrophic community of metazoans, mostly aquatic insect larvae (Beaver 1983; Lam et al. 2017), and microorganisms (Takeuchi et al. 2011). Finally, nutrients are absorbed by the pitcher through special glands, by active membrane transport (Moran et al. 2010) and endocytosis (Adlassnig et al. 2012), and the majority is quickly allocated to young growing tissues (Scharmann et al. 2013).

The evolutionary origin of *Nepenthes* lies in a group of non-core Caryophyllales (Cuénoud et al. 2002), for which carnivory was proposed to be an ancestral trait (Heubl et al. 2006; Renner & Specht 2011). Familiar non-carnivorous plants close to this group are Polygonaceae. *Nepenthes*' exact position is not clear but they are allied to Ancistrocladaceae, Dioncophyllaceae, Droseraceae (sundews, Venus flytrap), and Drosophyllaceae (Renner & Specht 2011). Strongly conflicting gene trees place *Nepenthes* partly as sister to Droseraceae and partly as sister to Drosophyllaceae–Ancistrocladaceae (Walker et al. 2017), implying an ancient rapid radiation or ancient introgression. The passive pitfall traps of *Nepenthes* may be derived from flat flypaper traps, which are produced by most of these allied carnivores, except for some species with rapidly moving snap-traps. However, conspicuous similarity at the level of morphology and molecular digestive machinery is observed between *Nepenthes* and the independently evolved pitchers of the carnivores Cephalotus (Oxalidales) and Sarraceniaceae (Ericales), amounting to a spectacular case of functional convergence (Juniper et al. 1989; Bauer et al. 2013; Fukushima et al. 2017).

While understanding of the basic physiology of *Nepenthes* traps has made large progress in recent years, much is still unclear about trap functional diversity and its evolution within the genus that contains more than 140 taxonomic species, distributed from Madagascar to New Caledonia, and from northeast India and southern China to northern Australia (Clarke 1997, 2001, Cheek & Jebb 2001, 2013, McPherson 2009, 2011). This evolutionary radiation has produced tremendous morphological variation of the pitcher trap, such that all its anatomical parts were modified, extended or eliminated in different combinations. Additional diversity exists at the level of prey spectra (e.g. Chin et al. 2014) and presumably in the digestive mechanisms, as the pitcher fluids range from slightly to extremely acidic, from watery to viscid (Bonhomme et al. 2011), from a few millilitres to more than one litre in volume, and contain different solute proteins (Biteau et al. 2013) as well as different communities of interacting animals (Beaver 1983; Kitching 2000) and microbes (Takeuchi et al. 2015; Sickel et al. 2016). Furthermore, this radiation has produced communities of multiple species that stably coexist in the same habitat but exhibit natural hybridisation. This striking diversity is still functionally unexplained to a large part. The factors promoting speciation and divergence in *Nepenthes* remain elusive, but could include adaptations in carnivory, among other more conventional drivers of plant speciation (Rieseberg & Willis 2007; Widmer et al. 2009; Givnish 2010).

*Nepenthes* also represents an independent case of dioecy in plants, as it is phylogenetically nested amidst entirely hermaphroditic allies (see above). No previous study has focused on this topic, but it is clear from taxonomic accounts that all species are dioecious, while sexual dimorphism is most prominent in the inflorescence. The mechanism of sex determination is unknown, although this knowledge is desirable for comparative evolutionary studies, and application in conservation and horticulture.

Previous evolutionary work on *Nepenthes* is dominated by molecular systematics, and mostly used subsets of species, small marker sets and phylogenetic approaches (Mullins 2000; Meimberg et al. 2001; Meimberg & Heubl 2006; Mokkamul et al. 2007; Alamsyah & Ito 2013; Suraninpong et al. 2015; Merckx et al. 2015; Bunawan et al. 2017). In spite of these efforts, many critical intrageneric relationships remain unclear. Consequently, *Nepenthes* phylogeography, the timescale of diversification, as well as the evolutionary history of traits (e.g. Bauer et al. 2012) remain challenging to interpret. A population genetic study has explored genetic geographic structure for a single species (Kurata et al. 2008). Comparative phylogenetic approaches of certain proteins used in prey digestion uncovered positive selection during the evolution of carnivory in the *Nepenthes* ancestors (e.g. Renner & Specht 2012), but it is unknown whether further adaptations occurred within the genus. Hence, the role of adaptation versus neutrality during the diversification of *Nepenthes* remains unresolved.

Neither experimental breeding nor quantitative genetics to understand the genetic architecture behind the carnivorous syndrome, and of sex determination have been attempted – probably because many species were only recently discovered, and *Nepenthes* plants have a long juvenile phase and can be challenging to cultivate. Despite this, a long tradition of horticultural breeding, starting in 19th century Victorian Britain, has established that many species combinations yield viable hybrid seed, and complex crosses are equally possible. Artificial hybrids are often vigorous and fertile in cultivation, and although failed attempts have typically not been reported, this suggests that reproductive isolation may be incomplete throughout the entire genus.

## Main methods

This project investigates evolutionary questions in non-model organisms mainly based on DNA and RNA sequences. Sequence data can illuminate evolutionary processes only in retrospective, and recourse to theoretical models is obligatory to make inferences about the underlying mechanisms. These theories are in general derived from fundamental biological processes (e.g. the principles of inheritance, the life cycle, population demography, etc.) and supported by empirical evidence (e.g. Hillis et al. 1992). The main frameworks I employed in this thesis are population genetics, in particular coalescent theory (Rosenberg & Nordborg 2002; Wakeley 2009) and to a lesser degree phylogenetics to study signatures of selection in coding sequences (e.g. Yang 2014) and the evolution of gene expression. In practice, all chapters are based on samples from natural *Nepenthes* populations and living plants that I collected in the field in Borneo (Brunei Darussalam) and Singapore, or plants that were artificially propagated in nurseries. The analysis of sequencing data involved bioinformatics and the development of computer code in Python, bash and R.

Acquisition of molecular sequence data for non-model organisms was until recently expensive and time consuming. Because only few genes or markers could

be sequenced or because of reliance on amplified fragment length polymorphisms, which only imperfectly reflect molecular sequences, the scope of many earlier projects and the questions that could be addressed were more limited. However, the rise of high throughput next generation sequencing (NGS) technology, spurred by biomedicine, has drastically reduced costs and speed of analyses and thus allowed for fundamentally broader, novel and more detailed perspectives on diverse patterns and processes in ecology and evolution. These technologies have themselves already undergone several major transitions and are continuously changing. Here I use various Illumina sequencing platforms in combination with current state-of-the-art bioinformatics tools.

In chapters I and II, I conduct surveys of population genetic patterns among multiple *Nepenthes* species. I genotyped these plants with double-digest restriction site associated DNA sequencing (ddRAD-seq, Peterson et al. 2012). It is a reliable library preparation technique that allows to study hundreds of thousands of DNA sites randomly distributed throughout genomes. Importantly, no prior genomic resources or marker development is required to apply this technique.

In chapters II and III, I use RNA-seq (Wang et al. 2009) to discover and quantify messenger-RNA transcripts from living tissues of cultivated *Nepenthes* plants. RNA-seq has replaced microarrays as the standard method for gene expression experiments, is easily adopted to non-model organisms, and is well supported by bioinformatic pipelines for assembly and quantification and by statistical tools for hypothesis testing. For chapter III, RNA-seq was combined with a feeding experiment to understand which genes are expressed during active feeding, and mass spectrometry proteomics for the de novo discovery of *Nepenthes* pitcher fluid proteins.

## Main objectives and research questions

This thesis investigated the evolutionary genomics of *Nepenthes* pitcher plants, with special attention to the role of selection and the carnivorous syndrome in their diversification, and the nature of their sex determination system. The following questions were addressed:

(1) How are multiple sympatric *Nepenthes* species related to each other, do they exchange genes under natural conditions, and which factors can explain their long-term ecological and genetic coexistence?

(2) Does dioecy in *Nepenthes* have a genetic basis, and what are the properties of this sex-determination system?

(3) Which are the genes that function in carnivory, and which selective regimes did they experience during the *Nepenthes* radiation? Were adaptive changes in genes with functions in carnivory more common than in genes with other functions?

## References

Abbott R, Albach D, Ansell S et al. (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229-246.

Adlassnig W, Koller-Peroutka M, Bauer S et al. (2012) Endocytotic uptake of nutrients in carnivorous plants. *The Plant Journal*, **71**, 303-313.

Akagi T, Henry IM, Tao R, Comai L (2014) A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science*, **346**, 646-650.

Alamsyah F, Ito M (2013) Phylogenetic analysis of Nepenthaceae, based on Internal Transcribed Spacer nuclear ribosomal DNA sequences. *APG Acta Phytotaxonomica et Geobotanica*, **64**, 113-126.

Athauda SBP, Inoue H, Iwamatsu A, Takahashi K (1998) Acid proteinase from *Nepenthes distillatoria* (Badura). In: *Aspartic Proteinases Advances in Experimental Medicine and Biology*. (ed James MNG), pp. 453-458. Springer US.

Athauda SBP, Matsumoto K, Rajapakshe S et al. (2004) Enzymic and structural characterization of Nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *Biochemical Journal*, **381**, 295-306.

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D (2015) The origins of reproductive isolation in plants. *New Phytologist*, **207**, 968-984.

Bachtrog D, Mank JE, Peichel CL et al. (2014) Sex determination: Why so many ways of doing it? *PLOS Biology*, **12**, e1001899.

Barrier M, Robichaux RH, Purugganan MD (2001) Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences*, **98**, 10208-10213.

Bauer U, Bohn HF, Federle W (2008) Harmless nectar source or deadly trap: *Nepenthes* pitchers are activated by rain, condensation and nectar. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 259-265.

Bauer U, Clemente CJ, Renner T, Federle W (2012) Form follows function: Morphological diversification and alternative trapping strategies in carnivorous *Nepenthes* pitcher plants. *Journal of Evolutionary Biology*, **25**, 90-102.

Bauer U, Federle W (2009) The insect-trapping rim of *Nepenthes* pitchers. Surface structure and function. *Plant Signaling & Behavior*, **4**, 1-5.

Bauer U, Paulin M, Robert D, Sutton GP (2015) Mechanism for rapid passive-dynamic prey capture in a pitcher plant. *Proceedings of the National Academy of Sciences*, **112**, 13384-13389.

Bauer U, Scharmann M, Skepper J, Federle W (2013) 'Insect aquaplaning' on a superhydrophilic hairy surface: How *Heliamphora nutans* Benth. pitcher plants capture prey. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122569.

Beaver RA (1983) The communities living in *Nepenthes* pitcher plants: fauna and food webs. In: *Phytotelmata: Terrestrial Plants as Hosts for Aquatic Insect Communities* (eds Frank JH, Lounibos LP), pp. 129-159. Plexus Publishing, Inc., Medford, New Jersey.

Bemm F, Becker D, Larisch C et al. (2016) Venus flytrap carnivorous lifestyle builds on herbivore defense strategies. *Genome Research*, **26**, 812-825.

Biteau F, Nisse E, Miguel S et al. (2013) A simple SDS-PAGE protein pattern from pitcher secretions as a new tool to distinguish *Nepenthes* species (Nepenthaceae). *American Journal of Botany*, **100**, 2478-2484.

Bohn HF, Federle W (2004) Insect aquaplaning: *Nepenthes* pitcher plants capture prey with the peristome, a fully wettable water-lubricated anisotropic surface. *Proceedings of the National Academy of Sciences*, **101**, 14138-14143.

Bonhomme V, Pelloux-Prayer H, Jousselin E et al. (2011) Slippery or sticky? Functional diversity in the trapping strategy of *Nepenthes* carnivorous plants. *New Phytologist*, **191**, 545-554.

Brawand D, Wagner CE, Li YI et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375-381.

Bunawan H, Yen CC, Yaakop S, Noor NM (2017) Phylogenetic inferences of *Nepenthes* species in Peninsular Malaysia revealed by chloroplast (trnL intron) and nuclear (ITS) DNA sequences. *BMC Research Notes*, **10**, 67.

Charlesworth B (1991) The evolution of sex chromosomes. *Science*, **251**, 1030-1033.

Charlesworth D (1999) Theories of the evolution of dioecy. In: *Gender and Sexual Dimorphism in Flowering Plants* (eds Geber MA, Dawson TE, Delph LF), pp. 33-60. Springer, Berlin, Heidelberg.

Charlesworth B, Charlesworth D (1978) A model for the evolution of dioecy and gynodioecy. *The American Naturalist*, **112**, 975-997.

Charnov EL (1982) The theory of sex allocation. In: *The Theory of Sex Allocation* Monographs in Population Biology., p. X+355P. Princeton University Press, Princeton, N.J., USA; Guilford, Surrey, England.

Cheek MR, Jebb MHP (2001) Nepenthaceae. *Flora Malesiana*, **15**, 1-157.

Cheek M, Jebb M (2013) The *Nepenthes micramphora* (Nepenthaceae) group, with two new species from Mindanao, Philippines. *Phytotaxa*, **151**, 25-34.

Chin L, Chung A, Clarke C (2014) Interspecific variation in prey capture behavior by co-occurring *Nepenthes* pitcher plants: Evidence for resource partitioning or sampling-scheme artifacts? *Plant Signaling & Behavior*, **9**, e27930.

Clarke CM (1997) *Nepenthes of Borneo*. Natural History Publications (Borneo), Kota Kinabalu.

Clarke CM (2001) *Nepenthes of Sumatra and Peninsular Malaysia*. Natural History Publications (Borneo), Kota Kinabalu.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Incorporated Publishers, Sunderland, MA, USA.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-3157.

Cuénoud P, Savolainen V, Chatrou LW et al. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *American Journal of Botany*, **89**, 132-144.

Darwin C (1859) *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. J. Murray, London.

Darwin C (1877) *The different forms of flowers on plants of the same species*. J. Murray, London.

Dieckmann U, Doebeli M, Metz JAJ, Tautz D (2004) *Adaptive Speciation*. Cambridge University Press.

Eilenberg H, Pnini-Cohen S, Schuster S, Movtchan A, Zilberstein A (2006) Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. *Journal of Experimental Botany*, **57**, 2775-2784.

Ellison AM (2006) Nutrient limitation and stoichiometry of carnivorous plants. *Plant Biology*, **8**, 740-747.

Fukushima K, Fang X, Alvarez-Ponce D et al. (2017) Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nature Ecology & Evolution*, **1**, 0059.

Gaume L, Forterre Y (2007) A viscoelastic deadly fluid in carnivorous pitcher plants. *PLOS ONE*, **2**, e1185.

Gaume L, Gorb S, Rowe N (2002) Function of epidermal surfaces in the trapping efficiency of *Nepenthes alata* pitchers. *New Phytologist*, **156**, 479-489.

Gavrilets S (2014) Models of speciation: where are we now? *Journal of Heredity*, **105**, 743-755.

Givnish TJ (2010) Ecology of plant speciation. *Taxon*, **59**, 1326-1366.

Givnish TJ (2015) Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, **207**, 297-303.

Givnish TJ, Burkhardt EL, Happel RE, Weintraub JD (1984) Carnivory in the bromeliad *Brocchinia reducta*, with a cost/benefit model for the general restriction of carnivorous plants to sunny, moist, nutrient-poor habitats. *The American Naturalist*, **124**, 479-497.

Gorb E, Haas K, Henrich A et al. (2005) Composite structure of the crystalline epicuticular wax layer of the slippery zone in the pitchers of the carnivorous plant *Nepenthes alata* and its effect on insect attachment. *Journal of Experimental Biology*, **208**, 4651-4662.

Gorb E, Kastner V, Peressadko A et al. (2004) Structure and properties of the glandular surface in the digestive zone of the pitcher in the carnivorous plant *Nepenthes ventrata* and its role in insect trapping and retention. *Journal of Experimental Biology*, **207**, 2947-2963.

Grant V (1981) *Plant Speciation*. Columbia University Press, New York, USA.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795-809.

Hatano N, Hamada T (2008) Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteome Research*, **7**, 809-816.

Hatano N, Hamada T (2012) Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteomics*, **75**, 4844-4852.

Heubl G, Bringmann G, Meimberg H (2006) Molecular phylogeny and character evolution of carnivorous plant families in Caryophyllales - revisited. *Plant Biology*, **8**, 821-830.

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ (1992) Experimental phylogenetics: generation of a known phylogeny. *Science*, **255**, 589-592.

Juniper BE, Robins RJ, Joel DM (1989) *The carnivorous plants*. Academic Press, London.

Käfer J, Marais GAB, Pannell JR (2017) On the rarity of dioecy in flowering plants. *Molecular Ecology*, **26**, 1225-1241.

Kapralov MV, Votintseva AA, Filatov DA (2013) Molecular adaptation during a rapid adaptive radiation. *Molecular Biology and Evolution*, **30**, 1051-1059.

Kimura M (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Kitching RL (2000) *Food webs and container habitats: The natural history and ecology of phytotelmata*. Cambridge University Press.

Kurata K, Jaffre T, Setoguchi H (2008) Genetic diversity and geographical structure of the pitcher plant *Nepenthes vieillardii* in New Caledonia: a chloroplast DNA haplotype analysis. *American Journal of Botany*, **95**, 1632-1644.

Lafon-Placette C, Vallejo-Marín M, Parisod C, Abbott RJ, Köhler C (2016) Current plant speciation research: unravelling the processes and mechanisms behind the evolution of reproductive isolation barriers. *New Phytologist*, **209**, 29-33.

Lam WN, Chong KY, Anand GS, Tan HTW (2017) Dipteran larvae and microbes facilitate nutrient sequestration in the *Nepenthes gracilis* pitcher plant host. *Biology Letters*, **13**, 20160928.

Lee L, Zhang Y, Ozar B, Sensen CW, Schriemer DC (2016) Carnivorous nutrition in pitcher plants (*Nepenthes* spp.) via an unusual complement of endogenous enzymes. *Journal of Proteome Research*, **15**, 3108-3117.

Losos JB, Mahler DL (2010) Adaptive radiation: The interaction of ecological opportunity, adaptation, and speciation. In: *Evolution Since Darwin: The First 150 Years.* (eds Bell MA, Futuyama DJ, Eanes WF, Levinton JS), pp. 381-420. Sinauer Assoc., Sunderland, MA, USA.

McPherson S (2009) *Pitcher plants of the Old World*. Redfern Natural History, Poole.

McPherson S (2011) *New Nepenthes - Volume 1*. Redfern Natural History, Poole.

Meimberg H, Heubl G (2006) Introduction of a nuclear marker for phylogenetic analysis of Nepenthaceae. *Plant Biology*, **8**, 831-840.

Meimberg H, Wistuba A, Dittrich P, Heubl G (2001) Molecular phylogeny of Nepenthaceae based on cladistic analysis of plastid trnK intron sequence data. *Plant Biology*, **3**, 164-175.

Merckx VSFT, Hendriks KP, Beentjes KK et al. (2015) Evolution of endemism on a young tropical mountain. *Nature*, **524**, 347-350.

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. *Annual Review of Plant Biology*, **62**, 485-514.

Mokkamul P, Chaveerach A, Sudmoon R, Tanee T (2007) Species identification and sex determination of the genus *Nepenthes* (Nepenthaceae). *Pakistan Journal of Biological Sciences*, **10**, 561-567.

Moore RC, Harkess AE, Weingartner LA (2016) How to be a seXY plant model: A holistic view of sex-chromosome research. *American Journal of Botany*, **103**, 1379-1382.

Moore JC, Pannell JR (2011) Sexual selection in plants. *Current Biology*, **21**, R176-R182.

Moran JA, Clarke CM (2010) The carnivorous syndrome in *Nepenthes* pitcher plants. *Plant Signaling & Behavior*, **5**, 644-648.

Moran JA, Clarke C, Gowen BE (2012) The use of light in prey capture by the tropical pitcher plant *Nepenthes aristolochioides*. *Plant Signaling & Behavior*, **7**, 957-960.

Moran JA, Hawkins BJ, Gowen BE, Robbins SL (2010) Ion fluxes across the pitcher walls of three Bornean *Nepenthes* pitcher plant species: flux rates and gland distribution patterns reflect nitrogen sequestration strategies. *Journal of Experimental Botany*, **61**, 1365-1374.

Moran JA, Moran AJ (1998) Foliar reflectance and vector analysis reveal nutrient stress in prey-deprived pitcher plants (*Nepenthes rafflesiana*). *International Journal of Plant Sciences*, **159**, 996-1001.

Mullins JT (2000) *Molecular Systematics of Nepenthaceae*. PhD Thesis. University of Reading, Reading, UK.

Nevado B, Atchison GW, Hughes CE, Filatov DA (2016) Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nature Communications*, **7**, 12384.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197-218.

Otto SP (2009) The evolutionary enigma of sex. *The American Naturalist*, **174**, 1-14.

Parker GA, Baker RR, Smith VGF (1972) The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology*, **36**, 529-553.

Pavlovič A, Saganová M (2015) A novel insight into the cost-benefit model for the evolution of botanical carnivory. *Annals of Botany*, **115**, 1075.

Pavlovič A, Singerova L, Demko V, Hudak J (2009) Feeding enhances photosynthetic efficiency in the carnivorous pitcher plant *Nepenthes talangensis*. *Annals of Botany*, **104**, 307-314.

Pavlovič A, Singerova L, Demko V, Šantrůček J, Hudak J (2010) Root nutrient uptake enhances photosynthetic assimilation in prey-deprived carnivorous pitcher plant *Nepenthes talangensis*. *Photosynthetica*, **48**, 227-233.

Pavlovič A, Slováková L, Šantrůček J (2011) Nutritional benefit from leaf litter utilization in the pitcher plant *Nepenthes ampullaria*. *Plant, Cell & Environment*, **34**, 1865-1873.

Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology*, **14**, e1002379.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE*, **7**, e37135.

Presgraves DC (2008) Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, **24**, 336-343.

Renner S, Ricklefs R (1995) Dioecy and its correlates in the flowering plants. *American Journal of Botany*, **82**, 596-606.

Renner T, Specht CD (2011) A sticky situation: Assessing adaptations for plant carnivory in the Caryophyllales by means of stochastic character mapping. International Journal of Plant Sciences, **172**, 889-901.

Renner T, Specht CD (2012) Molecular and functional evolution of class I chitinases for plant carnivory in the Caryophyllales. *Molecular Biology and Evolution*, **29**, 2971-2985.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences*, **99**, 12242-12245.

Rieseberg LH, Willis JH (2007) Plant Speciation. *Science*, **317**, 910-914.

Rodrigues N, Vuille Y, Loman J, Perrin N (2015) Sex-chromosome differentiation and 'sex races' in the common frog (*Rana temporaria*). *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142726.

Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, **13**, 505-516.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380-390.

Rottloff S, Miguel S, Biteau F et al. (2016) Proteome analysis of digestive fluids in *Nepenthes* pitchers. *Annals of Botany*, **117**, 479-495.

Roux J, Privman E, Moretti S et al. (2014) Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, **31**, 1661-1685.

Scharmann M, Thornham DG, Grafe TU, Federle W (2013) A novel type of nutritional ant-plant interaction: Ant partners of carnivorous pitcher plants prevent nutrient export by dipteran pitcher infauna. *PLOS ONE*, **8**, e63556.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York, USA.

Scholz I, Bueckins M, Dolge L et al. (2010) Slippery surfaces of pitcher plants: *Nepenthes* wax crystals minimize insect attachment via microscopic surface roughness. *Journal of Experimental Biology*, **213**, 1115-1125.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198-207.

Seehausen O, Butlin RK, Keller I et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176-192.

Sickel W, Grafe TU, Meuche I, Steffan-Dewenter I, Keller A (2016) Bacterial diversity and community structure in two Bornean *Nepenthes* species with differences in nitrogen acquisition strategies. *Microbial Ecology*, **71**, 1-16.

Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia University Press, New York, USA.

Suraninpong P, Nuanlaong S, Wuthisuthimethavee S (2015) A new classification of Thailands *Nepenthes* species by genetic analysis of AFLP markers. *Acta Horticulturae*, 1**100**, 77-82.

Takeuchi Y, Chaffron S, Salcher MM et al. (2015) Bacterial diversity and composition in the fluid of pitcher plants of the genus *Nepenthes*. *Systematic and Applied Microbiology*, **38**, 330-339.

Takeuchi Y, Salcher MM, Ushio M et al. (2011) In situ enzyme activity in the dissolved and particulate fraction of the fluid from four pitcher plant species of the genus *Nepenthes*. *PLOS ONE*, **6**, e25144.

The Marie Curie SPECIATION Network (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27-39.

Úbeda F, Patten MM, Wild G (2015) On the origin of sex chromosomes from meiotic drive. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20141932.

Wakeley J (2009) *Coalescent theory: an introduction*. Roberts & Company Publishers, Greenwood Village, Colorado, USA.

Walker JF, Yang Y, Moore MJ et al. (2017) Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *bioRxiv*, 115741.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57-63.

Westergaard M (1958) The mechanism of sex determination in dioecious flowering plants. In: *Advances in Genetics* (ed Demerec M), pp. 217-281. Academic Press.

Widmer A, Lexer C, Cozzolino S (2009) Evolution of reproductive isolation in plants. *Heredity*, **102**, 31-38.

Willson MF (1979) Sexual selection in plants. *The American Naturalist*, **113**, 777-790.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851-865.

Yang Z (2014) *Molecular Evolution: A Statistical Approach*. Oxford University Press, Oxford, United Kingdom; New York, NY, United States of America.

# Chapter 1
# Divergence with gene flow among multiple sympatric carnivorous *Nepenthes* pitcher plants is linked to trap morphology

a manuscript in revision co-authored by

M Scharmann[1], F Metali[2], T U Grafe[2], A Widmer[1]

[1] ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology (IBZ), Universitätstrasse 16, 8092 Zurich, Switzerland
[2] Faculty of Science, Universiti Brunei Darussalam, Jalan Tungku Link, Bandar Seri Begawan, BE1410 Brunei Darussalam

## Abstract

Speciation remains an incompletely understood phenomenon because it is difficult to establish the relative importance of neutral and adaptive processes during divergence, and what factors exert such selective pressures. In theory, long-term interspecific gene flow should either reverse speciation, or erode neutral drift-mutation divergence and thus expose adaptive divergence. Here we study mixed assemblages of carnivorous *Nepenthes* pitcher plants, inspired by frequently observed natural hybrids. Using genome-wide markers, we demonstrate that up to seven sympatric morphospecies of *Nepenthes* are distinguishable by population genetic and phylogenetic methods, yet exhibit clear signs of gene flow. Verified backcross hybrids confirm that reproductive isolation is incomplete. Comprehensive ABBA-BABA tests indicate pervasive genomic admixture in all species. Divergence with gene flow appears to have occurred continuously in a subset of species pairs, as inferred from coalescent simulations and Approximate Bayesian Computation. Finally, we focus on the shallow diverged sister species *N. hemsleyana* and *N. rafflesiana* t.f., and find that pitcher morphology, a key nutrient acquisition trait, shows diversifying selection (Pst–Fst comparison).

We suggest that adaptive divergence into multiple directions is accompanied by gene flow within sympatric *Nepenthes* assemblages, and that the morphology of carnivorous traps is involved. Furthermore, we provide adaptations of two powerful methods to test for introgression and possibly adaptive diversification in complexes of multiple species.

## Introduction

### Speciation and introgression

The evolution of reproductive isolation (RI) is central to the process of speciation. Before RI is complete, gene flow can challenge the integrity of incipient species. Hybridisation followed by repeated backcrossing can lead to introgression, the exchange of alleles between the gene pools of divergent genetic entities (Harrison & Larson 2014). Gene flow of sufficient strength will erode divergence over time, unless it is balanced by selection (Lenormand 2002). Consequently, the sympatric coexistence of related incipient species in the face of long-term realised gene flow (i.e. introgression) implies strong selection against at least some admixed genotypes (Lindtke et al. 2014; Christe et al. 2016). Else, the speciation process would stall or be reversed, leading to the loss of incipient species lacking complete RI (Vonlanthen et al. 2012). The scenario of ongoing divergence between incipient species in the face of long-term introgression is known as divergence with gene flow (DGF) or isolation with migration, and includes sympatric speciation, in which gene exchange occurs since onset of divergence, and secondary contact following a period of geographic isolation, as well as all sorts of situations in which gene flow happens intermittently (Pinho & Hey 2010). In addition to erasing neutral divergence between species, introgression can also be a source of adaptive variation (e.g. Racimo et al. 2015), and even promote divergence and speciation (e.g. homoploid hybrid speciation: Gross & Rieseberg 2005; hybrid swarm theory: Seehausen 2004). Speciation scenarios involving phases of incomplete RI and gene flow (e.g. Wu 2001) are in marked contrast to ’classic’ views (e.g. Mayr 1942) in which speciation proceeds only in the absence of gene flow, typically mediated by geographic isolation.

Long-term introgression among diverging taxa provides the exciting opportunity to study the role of selection in ecological divergence and speciation because the incipient species are not able to accumulate or retain strong divergence unrelated to RI (Schluter 2000). A common approach are outlier scans for interspecific differentiation, which aim to identify traits and genomic regions underlying RI and speciation, and frequently make the assumption that outliers result from resistance to homogenising gene flow. Through differential selection against migrant alleles, DGF can contribute to the formation of differentiation outlier loci and genomic islands of differentiation (Turner et al. 2005; Harrison 2012). However, outliers can also arise from processes unrelated to RI and speciation per se, including divergence after speciation due to independent adaptations in the descendant species (Cruickshank & Hahn 2014). To reject these alternative explanations in favour of diversifying selection it is necessary to verify the occurrence of interspecific gene flow.

### Detecting gene flow and introgression

Interspecific gene flow is difficult to detect in practice. While early-generation hybrids are relatively easy to identify, they do not necessarily imply realised

gene flow between the parental units because they could fail to reproduce and be genetic dead ends. Evolutionarily relevant long-term gene flow must be inferred from current patterns of genetic variation. However, similar patterns can result from distinct evolutionary processes. Two conspicuous observations that can indicate introgression are interspecific allele sharing and phylogenetic incongruence among different loci. However, the former may also reflect retained ancestral polymorphism and the latter can also result from random lineage sorting of ancestral polymorphisms (Degnan & Rosenberg 2009). It is thus clearly necessary to specifically consider these confounding processes when testing for gene flow.

Most studies of introgression investigated species pairs with geographically defined hybrid zones (Barton & Hewitt 1985; Anderson & Thompson 2002; Buerkle 2005; Gompert & Buerkle 2011). The hybrid zone framework offers great conceptual advantages to detect diversifying selection, but it is not useful in settings with multiple species and uncertain or nested distribution ranges. Consequently, genomic studies testing introgression among multiple strictly sympatric species are rare. Only 12 out of 42 studies reviewed by Payseur & Rieseberg (2016) investigated more than two at least partially sympatric species. Recent studies of groups that might involve ongoing gene flow between more than two sympatric species include lake-specific cichlid radiations (Meyer et al. 2016; Meier et al. 2017), North American *Quercus* (Eaton et al. 2015), *Pedicularis* (Eaton & Ree 2013), *Heliconius* butterflies (The Heliconius Genome Consortium 2012; Martin et al. 2013), *Anopheles* spp. (Fontaine et al. 2015), and Darwin's finches (Lamichhaney et al. 2015). One explanation for this relative shortage of examples may be a lack of appropriate methods to study introgressing sympatric species groups, although they might be common in nature.

Here we adapt two methods to test for introgression in sympatric assemblages of closely related species. Patterson's D-test or ABBA-BABA statistic (Durand et al. 2011) quantifies deviation from a bifurcating phylogenetic species tree, but introgression can be identified in only one pair of species in its conventional application. The combination of coalescent theory (Rosenberg & Nordborg 2002; Wakeley 2009) with simulations and Approximate Bayesian Computation (ABC: Tavare et al. 1997; Beaumont et al. 2002) is a particularly powerful tool to distinguish between gene flow and alternative evolutionary processes, and to estimate gene flow parameters. However, processes critical to DGF were not implemented until recently (Roux et al. 2013, 2014), the handling of more than a few hundred markers was impractical, and the presence of more than two species could confound results.

***Nepenthes* pitcher plants**

Our study system are sympatric communities of carnivorous *Nepenthes* pitcher plants, in which phenotypic observations suggested that introgression might be an important evolutionary process. *Nepenthes* (Nepenthaceae, Caryophyllales) contains over 140 taxa of perennial carnivorous vines and shrubs (Cheek & Jebb 2001,

2013). Botanical carnivory in general is understood as a nutrient sequestration strategy, supplying minerals lacking from soils (Ellison 2006). *Nepenthes* plants capture, drown and digest animals in their pitcher traps, which are highly modified leaves with complex biomechanical and physiological adaptations (Juniper et al. 1989; Moran & Clarke 2010). Within the genus, there is limited variation in flowers, inflorescence architecture and vegetative traits, but tremendous variation on the common theme of the pitcher traps regarding size, shape, colouration, life span, and multiple other characters. However, no satisfactory explanation has yet been put forth for this diversity that was proposed to represent an adaptive radiation (Meimberg & Heubl 2006; Bauer et al. 2012; Pavlovič 2012). One hypothesis is that carnivory itself promoted diversification through competition for prey, pushing coexisting species to exploit different prey spectra (Moran et al. 1999; Bauer et al. 2011; Chin et al. 2014; Gaume et al. 2016). Implicitly, this hypothesis predicts diversifying selection acting on trap morphology.

Here we aim to elucidate this prediction by exploiting the theoretical expectation that in migration-selection equilibrium, genes and traits under diversifying selection will continue to diverge, while neutral genes and traits move freely between species. In other words, we hypothesise that long-term interspecific gene flow among sympatric *Nepenthes* could indirectly reveal diversifying selection, which might act on pitcher trap morphology. First generation natural hybrids among sympatric *Nepenthes* are well documented at the phenotypic level (Cheek & Jebb 2001; Clarke 2001; Peng & Clarke 2015) and motivated this study. They verify that reproductive isolation (RI) is incomplete. However, there are only few and untested records of later generation hybrids in the wild, and no data on hybrid reproductive success is available.

We focussed on the detection of long-term gene flow (i.e. DGF) in the particularly rich *Nepenthes* assemblage in the lowland peat swamp and heath forests of Borneo, where seven or more species may grow within a few hundred meters distance (Figure 1). Three of these species were also studied in Singapore. Specifically, we addressed the following questions:

(1) Are *Nepenthes* morphospecies distinct population genetic entities, and how are they related to each other?

(2) What is the evidence for ancient introgression and ongoing divergence with gene flow between them?

(3) Is divergence in *Nepenthes* pitcher trap morphology higher than expected under neutral drift-mutation evolution?

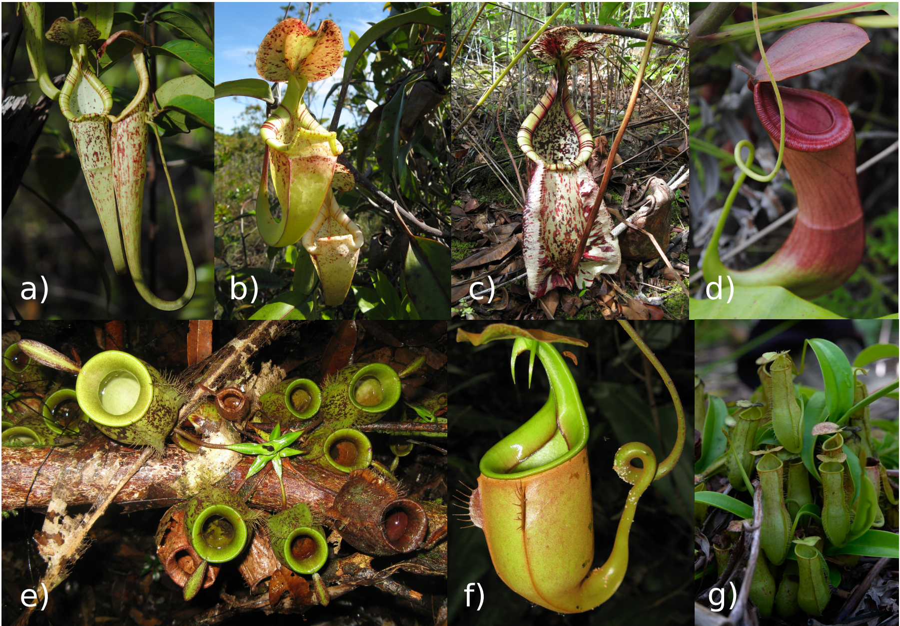



**Figure 1.** Portraits of the sympatric morphospecies sampled in Brunei. a) *Nepenthes hemsleyana*, upper pitchers; b) *N. rafflesiana* typical form (t.f., Borneo endemic), upper pitchers; c) *N. rafflesiana* giant form (g.f., Borneo endemic), lower pitcher; d) *N. mirabilis*, upper pitcher (depicted is the 'typical variety whereas in Brunei this species is represented by the variety *echinostoma*) e) *N. ampullaria*, rosette with lower pitchers; f) *N. bicalcarata*, upper pitcher; g) *N. gracilis*, lower pitchers. Photos: M. Scharmann.

## Materials and Methods

### Sampling design and field sites

We sampled natural populations of sympatric *Nepenthes* in Brunei Darussalam, Borneo (c. 4°36'N 114°35'E) and Singapore (c. 1°21'N 103°48'E) between July and September 2013. We sampled between 10 (minimum 6) and 20 individuals from each of the three to seven locally most abundant *Nepenthes* species (Figure 1): *N. ampullaria* (Borneo and Singapore), *N. bicalcarata* (Borneo), *N. mirabilis* (Borneo: represented by *N. mirabilis* var. *echinostoma*), *N. gracilis* (Borneo and Singapore), *N. rafflesiana* (Singapore), and *N. hemsleyana* (Borneo). Morphospecies determination was based on current taxonomy (Cheek & Jebb 2001; Scharmann & Grafe 2013). In addition, we distinguished in Borneo among two "forms" of *N. rafflesiana*, i.e. the "typical form" (*N. rafflesiana* t.f.) and the "giant form" (*N. rafflesiana* g.f.), as informally described by Clarke (1992, 1997). *N. albomarginata* and *N. hispida* also occur in Bruneian lowlands, but their rarity made them unsuitable for population sampling and less likely to contribute to introgression.

We sampled plants in a transect-like design with a minimum distance between conspecific individuals of 50 m to avoid potential small-scale genetic structure and clonality. Most samples were taken in primary peat swamp or heath forest. In Singapore, only secondary vegetation was available. All species were sampled from mixed stands along the same transects. Plants were well established, large and climbing or with extensive rosettes. To serve as a non-introgressing outgroup, we included samples of *N. pervillei* collected on the island of Mahé, Seychelles. Leaf material was stored in a nucleic acid preserving buffer solution (Camacho-Sanchez et al. 2013).

**ddRAD-seq**

DNA was extracted using silica-column kits (Nucleospin Plant II, Macherey Nagel, Düren, Germany). Initial problems with low yield and DNA fragmentation were overcome by grinding the woody leaves to fine powder cooled by liquid nitrogen. Libraries were produced following a ddRAD protocol (Peterson et al. 2012) with slight modifications. In brief, genomic DNA was restriction-digested with two enzymes (EcoRI and TaqI), and a 5-base barcode annealed to the EcoRI overhang. Illumina adaptors with indices were added to the TaqI overhang to allow further multiplexing. By selective retention of fragments with restriction overhangs and a size selection step, only a minor but relatively consistent part of the genomes is ultimately incorporated into the library. Sequencing of ddRAD libraries was undertaken in pools of 42 samples each, and each pool was sequenced for 101 bp single-end reads in half an Illumina HiSeq lane. Genotyping followed a dDocent pipeline (Puritz et al. 2014) modified for single-end reads, and specific genotype-missingness filters were applied in each analysis. Full bioinformatics details are provided in Supporting Information S1.

**Genetic population structure and hybrid identification**

Correspondence between morphospecies and genetic groups was tested per location. Sites were filtered to be variable and present in at least 90% of samples per location. Principal Component Analyses were conducted with the tool smartpca (EIGENSTRAT package, Price et al. 2006), and plotted in R (R Core Team 2014). Neighbour-joining trees were built from pairwise Kimura distances using rapidNJ (Simonsen & Pedersen 2011). Hybrids were identified using fastSTRUCTURE (Raj et al. 2014) with a simple prior by considering all pairwise comparisons of morphospecies in separate runs, enforcing k=2 each time. Nucleotide sites were filtered to be present in at least 80% of samples per species, and sites invariant in a given morphospecies pair were excluded. Hybrids were defined as individuals assigned with >1% probability to another fastSTRUCTURE cluster than the one representing the a priori assigned morphospecies. All hybrids were excluded from further analyses.

**Phylogenetic analysis**

We reconstructed bifurcating phylogenetic ”backbone” topologies among all remaining individuals and populations for subsequent introgression tests. Sites

were filtered to be present in at least 20 out of the total 140 individuals, plus 20 individuals of *N. pervillei* from the Seychelles as an outgroup. Genotypes were then converted to a concatenated supermatrix of SNPs only, retaining heterozygotes as ambiguity code. This alignment contained 1,373,129 sites with 64.3% gap characters. High missingness does not confound tree topology when it accompanies an increased number of sites (Leache et al. 2015). A maximum likelihood tree was estimated under the GTRCAT substitution model using ExaML v.3.0.17, with a starting parsimony tree generated in RAxML v8.2.4 (Stamatakis 2014). Tree uncertainty was quantified by the Shimodaira-Hasegawa-like approximate likelihood ratio test. This indicator was chosen instead of RAxMLs rapid bootstrap algorithm because of superior speed and performance (Simmons & Norton 2014). A population-level tree was created by collapsing individuals of monophyletic populations into single tips.

**Pairwise interpretation of ABBA-BABA statistics**

We tested for introgression against this backbone tree using Patterson's D-tests (ABBA-BABA tests, Durand et al. 2011). We reported the $f_d$ statistic (Martin et al. 2015) instead of Patterson's D because of lower variance and a linear relationship with the amount of introgression. The $f_d$ statistic does not overestimate introgression and can be understood as the proportion of the genomes shared through introgression.

Patterson's D-test is based on four populations and can reveal introgression only among at most two of them. In this study we aimed to summarise multiple Patterson's D-tests, which would allow massive 'pairwise' testing for introgression signals in arbitrarily large bifurcating trees, a strategy that was already suggested but not yet explored (e.g. Pease & Hahn 2015). To this end, we interpreted each D-test as two pairwise tests: We reasoned that the single test on (((p1,p2),p3),outgroup) is actually informative for two pairs of populations. It quantifies derived allele sharing in the pair p1–p3 as contrasted to the pair p2–p3. The sign of the D-statistic only indicates which of the two pairs shares more derived alleles (ABBA or BABA excess), and the absolute value of the D-statistic describes that pair. The remaining pair shows no evidence for excessive derived allele sharing. For example, a BABA-excess of D=-0.2 for the test (((p1,p2),p3),outgroup) conveys that D(p1–p3)=0.2 and also that D(p2–p3)=0. Note that D=0 does not indicate a lack of derived allele sharing and introgression but rather a lack of statistical power due to the contrast pair outweighing the level of derived allele sharing.

We found all possible combinations (triplets) that conform to the topology (((p1,p2),p3), *N. pervillei*) by iteratively pruning the complete maximum-likelihood trees down to four tips. *N. pervillei* was chosen as the outgroup in all tests because it has a basal position within *Nepenthes* (Mullins 2000; Meimberg et al. 2001) and is endemic to the Seychelles, hence not introgressing with the Asian species. ABBA-BABA patterns were analysed at the population level, testing all 120 possible triplets, while at the individual level, we tested only a subset

of 86,328 out of 447,580 possible triplets (19.3%), by randomly down-sampling from the possible triplets, such that each pair of individuals (pairs p1–p3 resp. p2–p3) occurred in at most ten triplets. Multiple ABBA-BABA tests conducted for a backbone tree were summarised and plotted in R as a pairwise matrix of f_d, retaining only the maximum f_d for each pair. Further details are provided in Supporting Information S2.

**Coalescent modelling and ABC**

Coalescent simulation with migration parameters and ABC for more than three populations is extremely difficult due to (1) enormous spaces to be explored (branching topology, model parameters), (2) lack of mathematical basis and computational implementation (joint summary statistics for more than two or three populations, efficient code), and (3) low number of RAD-tags shared among all populations. Currently, the only feasible approach to resolve introgression among many populations is to test pairwise combinations. We assume in this study that the history of introgression between any pair of species can be represented by either of two highly simplified models, and aim to classify *Nepenthes* species pairs accordingly. In the "isolation" model (Figure 2 A), an ancestral species instantly splits into two perfectly isolated descendants. The divergence with gene flow model ("DGF", Figure 2 B) follows the same course, but the descendant species exchange migrants at a constant rate since their split and continuously up to the present. Migration may be asymmetric and heterogeneous among loci. We acknowledge that real speciation histories may be more complex (e.g. different phases of gene flow, Roux et al. 2013; Christe et al. 2017), but the aim of this study was to detect tendencies towards one or the other extreme of speciation modes – "strict isolation" or "divergence with continuous gene flow".

However, naïve reduction of assemblages of multiple potentially introgressing species to pairwise analyses may be confounded because apparent migration signals in any focal pair may actually be migration from the excluded species. Unobserved species (populations) that appear only as statistical signals are called "ghosts" (Beerli 2004; Slatkin 2005). To avoid the confusion of "DGF" in focal species pairs with migration from ghosts, we considered them explicitly in the "ghost model" (Figure 2 C). This is essentially like the "isolation" model but in addition there is migration from a ghost species into both descendants. A single ghost in the model can represent multiple actual taxa through a large ghost population size. We further constrained ghosts to be donors but never recipients of migrants because this would constitute a "DGF" model through indirect migration between the focal species. In a bifurcation model that describes the evolution of two focal species, three distinct ghosts are possible. These are the three different topologies by which any two species can be related to a third: an external ghost branches off from the overall common ancestor before the focal species diverge, and two different internal ghosts are sisters to either one of the focal species.

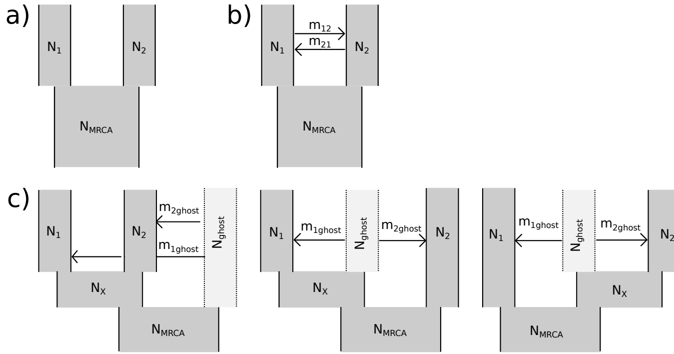



**Figure 2.** Schemes of speciation models explored in this study. At the top of each diagram is the present time, when sequences from two distinct species (grey bars, effective population size N) are sampled. Sequences find common ancestors going backwards into the past towards the bottom of the diagram, until they are merged into an ancestral species where they finally coalesce into a single ancestor. a) isolation model. b) divergence with gene flow (DGF) model: migration rates possibly asymmetric and heterogenous among loci. b) ghost model: no migration between observed species, but both species receive migrants from an unobserved ghost species (lighter grey). From left to right: external ghost, internal ghost sister to species 1, internal ghost sister to species 2.

The general design of our simulation and ABC pipeline closely follows the approach pioneered by Roux et al. (2013) but was amended and re-written. Free model parameters were the population sizes, migration rates, and split time. We implemented heterogeneity in migration rates among loci, a conceptual implication of introgression and an important diagnostic property (Roux et al. 2013, 2014), as a binomial probability, i.e. the proportion of loci having a migration rate greater than zero and of the magnitude given by a separate migration rate parameter. For illustration, these two parameters were multiplied, yielding a genome-wide average migration rate. We used a total of 114 summary statistics, including several that are novel to ABC, such as the joint site frequency spectrum resp. allele frequency spectrum (Gutenkunst et al. 2009). The sampling scheme of the real data was reproduced as closely as possible in simulations, i.e. the same numbers of samples and RAD-tags, and the same sequence lengths. For each of the 24 species pairs in Brunei and Singapore, we retained 1,419-12,349 RAD-tags (filters: minimum read depth six, presence in at least 80% of individuals per species, excluding hybrids, excluding RAD-tags with within-species heterozygote excess, retaining monomorphic RAD-tags). Observed data were then classified into the three considered models (Figure 2). ABC model choice was conducted with the Random Forest algorithm (R package abcrf, Pudlo et al. 2016), and parameters were estimated using neural networks (R package abc, Csilléry et al. 2012). Further details and a flowchart are presented in Supporting Information S3.

### Pst–Fst comparison

Phenotypic data for *N. hemsleyana* and *N. rafflesiana* t.f. were taken from Lim et al. (2015) who sampled the same populations as this study. We tested whether phenotypic divergence (Pst) between these sister species deviated from expectations for quantitative traits under neutrality (Leinonen et al. 2013). Pst is analogous to Qst but is estimated without a breeding experiment. We estimated Pst = ( $c/h^2 \times$ variance_between_species ) / ( $c/h^2 \times$ variance_between_species + $2 \times$ variance_within_species ), where the variance components are the residuals from a linear regression model (R: lm), $h^2$ is the overall proportion of variance from additive genetic effects (heritability), and c is the proportion of variance from additive genetic effects across species (Brommer 2011). As the phenotypes came from a natural common garden (Brunei), we assumed $c=h^2=1$, and addressed the robustness of this assumption by calculating the critical $c/h^2$ ratio (Brommer 2011). This is the $c/h^2$ ratio above which the bootstrapped 95% confidence intervals of Fst and Pst do not overlap anymore, hence lower critical $c/h^2$ ratios indicate stronger robustness of the Pst–Fst comparison. Fst for the population pair *N. hemsleyana* (n=17) and *N. rafflesiana* t.f. (n=18) was estimated following Weir & Cockerham (1984). Hybrids were excluded and individuals and genotypes filtered as in the ABC approach but monomorphic RAD-tags were removed and negative Fst estimates set to zero.

The difference between Pst and Fst was calculated using a custom bootstrap procedure in R (bootstrap resampling from the observed Pst distribution (n = 94) and observed Fst distribution (n = 10,397) at equal sample sizes for both (n = 94), then calculation of the difference in means; 10k replicates). Trait neutrality was rejected if the bootstrapped Pst–Fst distribution excluded zero, and if the critical $c/h^2$ ratio was below 0.2.

## Results

### Genetic population structure and hybrids

In both Brunei and Singapore, we found strong agreement between morphologically determined groups and single genetic clusters (Neighbour-Joining trees and Principal Component Analyses, Figures in Supporting Information S4). However, several phenotypically inconspicuous individuals fell in between the main clusters in NJ trees and PCAs. We tested the hypothesis that these were hybrids using individual ancestry proportions (Figures Supporting Information S5 and S6). Apparent ancestry from more than two species may be explained by genetic similarity among the potential ancestors rather than complex hybrid origin –in such cases the two species with the highest assignment probabilities were accepted as the true ancestors to satisfy parsimony. We identified 12 hybrids (Table 1) and all other individuals were assigned to a single genetic cluster with >99% probability. In general, only few samples per population were hybrids. *N. rafflesiana* g.f. was an exception because six out of 12 samples initially identified as this morphospecies were hybrids. All individuals identified as hybrids were

excluded from further analyses, because we aimed to study natural long-term introgression. The relatively recent formation of these hybrid individuals may be linked to human activity in or close to our sampling locations (deforestation), so we cannot exclude that non-natural causes underlie their occurrence.

**Table 1.** *Nepenthes* hybrids found in Brunei, detected using thousands of SNPs and fastSTRUCTURE. Back-cross is here a category for any plant that did not show close to 50% membership in two clusters.

| morphospecies | fastSTRUCTURE assignment probabilities >1% to non-self cluster | conclusion | habitat |
|---|---|---|---|
| *ampullaria* | 30% *gracilis* | backcross *ampullaria–gracilis* | border to primary forest |
| *rafflesiana* g.f. | 59% *hemsleyana*, 3% *raff.* t.f. | backcross *hemsleyana–raff.* g.f. | previously logged forest |
| *rafflesiana* g.f. | 70% *hemsleyana*, 61% *raff.* t.f. | backcross *raff.* g.f.–*raff.* t.f. / complex w. *hemsleyana* | secondary vegetation |
| *rafflesiana* g.f. | 19% *raff.* t.f., 2% *hemsleyana* | backcross *raff.* g.f.–*raff.* t.f. | secondary vegetation |
| *rafflesiana* g.f. | 26% *raff.* t.f., 19% *hemsleyana* | backcross *raff.* g.f.–*raff.* t.f. | secondary vegetation |
| *rafflesiana* g.f. | 40% *raff.* t.f., 14% *hemsleyana* | F1 or backcross *raff.* g.f.–*raff.* t.f. | secondary vegetation |
| *rafflesiana* g.f. | 53% *raff.* t.f., 34% *hemsleyana* | F1 or backcross *raff.* g.f.–*raff.* t.f. | secondary vegetation |
| *hemsleyana* | 14% *raff.* t.f | backcross *hemsleyana–raff.* t.f. | degraded kerangas soil |
| *hemsleyana* | 14% *raff.* t.f. | backcross *hemsleyana–raff.* t.f. | degraded kerangas soil |
| *hemsleyana* | 46% *mirabilis*, 9% *ampullaria*, 7% *bicalcarata*, 5% *gracilis*,(72% *raff.* g.f.) | F1 *hemsleyana–mirabilis* | secondary vegetation |
| *mirabilis* | 32% *ampullaria*, 5% *hemsleyana*, 5% *raff.* t.f., 4% *raff.* g.f. | backcross *mirabilis–ampullaria* | secondary vegetation |
| *rafflesiana* t.f. | 10% *hemsleyana*, 4% *raff.* g.f. | backcross *hemsleyana–raff.* t.f. | degraded kerangas soil |

### Population phylogeny and ABBA-BABA statistics

A maximum-likelihood tree excluding hybrids suggested that the sampled *Nepenthes* populations and morphospecies were monophyletic groups (summarised

to the population level in Figure 3 left, full individual-based tree Figure Supporting Information S7-1 left). With the distantly allopatric *N. pervillei* (Seychelles) as outgroup, a basal dichotomy separated the sister species *N. bicalcarata* and *N. gracilis* (weakest support) from a branch containing populations of *N. ampullaria*, *N. mirabilis* and the *N. rafflesiana* group. Within this group, *N. rafflesiana* Singapore appeared basal while the Bornean species *N. rafflesiana* g.f., *N. rafflesiana* t.f. and *N. hemsleyana* formed a clade.

We then tested for signatures of introgression against this bifurcating backbone tree using the same genotypes and multiple Patterson's D-tests (ABBA-BABA tests). All *Nepenthes* contained signals of introgression with other populations, sharing at least 2.7-16.3% of the genomes in excess over incomplete lineage sorting against the bifurcation hypothesis (pairwise f_d matrix, Figure 3 right). The fewest introgression signals were recorded with *N. bicalcarata*, whereas its putative sister species *N. gracilis* introgressed with all other populations, especially with *N. ampullaria*. Introgression appeared even more pervasive when ABBA-BABA tests were based not on population allele frequencies but on individual genotypes, as in the original human–Neanderthal test (Green et al. 2010; pairwise f_d matrix, Figure Supporting Information S7-1 right). Although interspecific f_d varied among individuals within populations, we detected highly significant f_d ($\geq 500$ informative sites, $p \leq 0.0001$) of up to 45% in all pairs of *Nepenthes* morphospecies/populations except *N. gracilis*–*N. bicalcarata*, *N. mirabilis*–*N. rafflesiana* g.f., and *N. mirabilis*–*N. hemsleyana*. Prominent patterns included an introgression "block" between most *N. mirabilis* and *N. rafflesiana* Singapore, and more introgression with the lineages *N. ampullaria*–*N. mirabilis*–*N. rafflesiana* for *N. gracilis* Singapore than for *N. gracilis* Brunei.

**Coalescent model choice and parameter estimation with Approximate Bayesian Computation**

After establishing the major phylogenetic relationships and assessing introgression, we assessed DGF in the framework of coalescent modelling. In brief, we tested whether bifurcating speciation models with or without continuous gene flow can explain the empirical results, and then estimated model parameters, including the strength and direction of gene flow.

The observed population genetic patterns for 24 sympatric *Nepenthes* species pairs (Supporting Information S8) covered a broad range of evolutionary divergence, illustrated by the low extreme *N. rafflesiana* t.f.–*N. hemsleyana* in Brunei with genome-wide mean net divergence of 0.1% (Nei & Li 1979) and Fst=0.11, and the high extreme *N. bicalcarata*–*N. gracilis* in Brunei with net divergence of 1.5% and mean Fst=0.63. Notably, the genome-wide Fst distributions were bimodal –with distinct peaks close to zero and one separated by a plateau –in all except the four least diverged species pairs, and all 24 joint site-frequency spectra contained SNPs segregating in both species (Figures Supporting Information S9). Sympatric species pairs were then formally classified by ABC into the models "DGF", "isolation" and "ghost". The latter was included because each *Nepenthes*

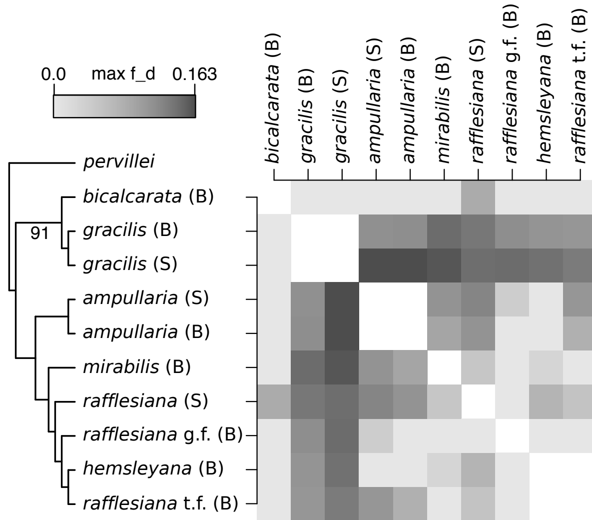



**Figure 3.** Phylogenetic relationships among studied *Nepenthes* populations (left, maximum-likelihood tree, branch lengths not shown, pruned from individual-based tree, B = Brunei, S = Singapore) and pairwise introgression matrix (right, maximum f_d statistics from multiple ABBA-BABA tests). All branches except one (annotated) had full SH-like aLRT support. Each f_d value greater 0.0 (lightest grey) resulted from a significant ABBA-BABA test ($p \leq 0.05$) based on 1,216-8,158 informative sites, each present in at least 50% of samples per population. Blank cells indicate impossible tests.

community consisted of more than two species, and gene flow from third parties (extra-pair species or "ghosts") can mimic the signature of "DGF" in a pair of species. This concern was substantiated by the classification error results from the Random Forest training phase: confusion of "ghost" as "DGF" occurred for 6-8% of simulations, while confusion of "isolation" as "DGF" occurred for only 0.7-1% of simulations.

Out of a total of 24 interspecific population pairs, "DGF" was rejected for nine pairs, but explained the data well for 15 pairs, and was the only appropriate speciation model for six of these pairs (Table 2). "Isolation" described the data for nine pairs well but all of these were also consistent with "ghost" and some even with "DGF". Results from these multiple pairwise comparisons indicate that introgression is prevalent among sympatric *Nepenthes* in Brunei, which can be summarised as an interaction network (Figure 4). At one extreme, *N. ampullaria* exhibited signatures of DGF with each of the other species. But even the least connected species, *N. gracilis*, showed DGF with two other species. In Singapore, however, the "DGF" scenario was rejected in favour of "ghost" and "isolation" for all three species pairs (Table 2).

.

**Table 2.** Results of coalescent modelling and ABC for all sympatric *Nepenthes* species pairs at two sampling locations, Brunei and Singapore. Speciation models that can explain the observed data are marked by an X – at least one of "DGF","ghost" and "isolation". Model parameters (genome-wide averages over all RAD-tags, medians of the joint posterior distributions) were estimated for "DGF" whenever it was not rejected, and otherwise for "isolation". Parameter units: N1, N2, Nanc: number of gene copies estimated in species 1, 2 and the ancestral population; Tsplit: generations ago; m12, m21: migration rates in $2 \times N \times m$, m12 = migration into 1 from 2. m21 = migration into 2 from 1. All units scaled by spontaneous mutation rate $\mu = 2.5 \times 10^{-8}$ per site per generation. Detailed statistics in Supporting Information S11.

| location | species 1 | species 2 | DGF | ghost | isolation | N1 | N2 | Nanc | Tsplit | m12 | m21 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Brunei | *ampullaria* | *bicalcarata* | X | X | X | 24500 | 4790 | 139600 | 167760 | 1.3 | 1.22 |
| | *ampullaria* | *raff.* g.f. | X | X | | 38120 | 14400 | 129750 | 23840 | 0.05 | 2.37 |
| | *ampullaria* | *gracilis* | X | X | X | 29040 | 7770 | 162640 | 142640 | 1.21 | 0.17 |
| | *ampullaria* | *hemsleyana* | X | | | 16040 | 46140 | 144250 | 75440 | 3.68 | 0.2 |
| | *ampullaria* | *mirabilis* | X | X | X | 20960 | 24530 | 134990 | 23800 | 0.98 | 0.03 |
| | *ampullaria* | *raff.* t.f. | X | | | 28620 | 18330 | 109430 | 73000 | 1.52 | 0.1 |
| | *bicalcarata* | *raff.* g.f. | | X | | 16760 | 32150 | 128510 | 63560 | | |
| | *bicalcarata* | *gracilis* | X | X | | 38320 | 23540 | 148250 | 463160 | 0.51 | 0.26 |
| | *bicalcarata* | *hemsleyana* | X | X | | 4360 | 28490 | 77240 | 102720 | 1.29 | 0.62 |
| | *bicalcarata* | *mirabilis* | X | X | X | 4070 | 23690 | 144900 | 59600 | 1.35 | 0.64 |
| | *bicalcarata* | *raff.* t.f. | | X | | 13210 | 23850 | 118680 | 75040 | | |
| | *raff.* g.f. | *gracilis* | | X | | 16800 | 6330 | 163040 | 38040 | | |
| | *raff.* g.f. | *hemsleyana* | X | | | 23390 | 39150 | 98280 | 26520 | 1.82 | 0.06 |
| | *raff.* g.f. | *mirabilis* | X | X | | 19170 | 42220 | 64510 | 33560 | 1.36 | 0.03 |
| | *raff.* g.f. | *raff.* t.f. | X | | | 27210 | 30820 | 125660 | 126600 | 5.96 | 2.91 |
| | *gracilis* | *hemsleyana* | | X | | 17220 | 35000 | 116440 | 92120 | | |
| | *gracilis* | *mirabilis* | | X | X | 10180 | 15310 | 123220 | 44120 | | |
| | *gracilis* | *raff.* t.f. | | X | | 23210 | 16790 | 119200 | 88320 | | |
| | *hemsleyana* | *mirabilis* | X | | | 35690 | 32220 | 114110 | 53080 | 0.58 | 0.27 |
| | *hemsleyana* | *raff.* t.f. | X | X | X | 28850 | 10060 | 155400 | 150880 | 22.75 | 7.04 |
| | *mirabilis* | *raff.* t.f. | X | | | 20150 | 25560 | 129950 | 115360 | 1.33 | 0.54 |
| Singapore | *ampullaria* | *gracilis* | | X | X | 24550 | 20650 | 109410 | 68080 | | |
| | *ampullaria* | *rafflesiana* | | X | X | 16010 | 32380 | 98220 | 56760 | | |
| | *gracilis* | *rafflesiana* | | X | X | 16860 | 24380 | 89310 | 68320 | | |

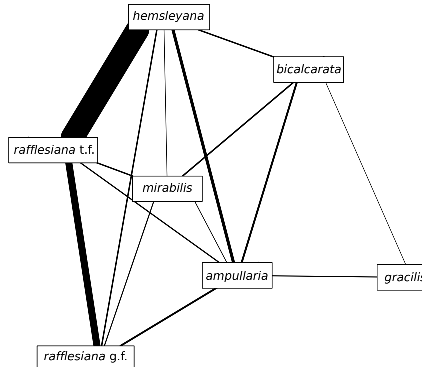



**Figure 4.** Interaction network for the *Nepenthes* community in Brunei, indicating introgression as estimated from coalescent models. Line width indicates strength of cumulative bi-directional gene flow. This illustration summarises results from pairwise models and does not represent a formally tested model encompassing all species.

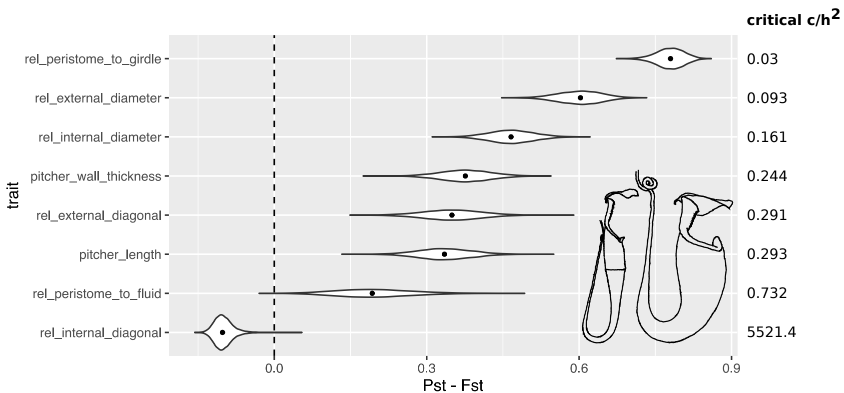



**Figure 5.** Comparison of observed phenotypic divergence (Pst) against neutral divergence (Fst, genome-wide mean = 0.11) in several upper pitcher trap traits for the direct sister species *N. hemsleyana* and *N. rafflesiana* t.f. in Brunei. Shown are the full distributions (violins, means as dots) of the Pst–Fst difference (X-axis) for 10,000 bootstrapped combinations of phenotypes and genetic markers for $c/h^2 = 1$, and the critical $c/h^2$ above which the 95% confidence intervals of Pst and Fst do not overlap. Insets in the lower right corner illustrate the divergent upper pitcher shapes (lateral, scaled to same length) of *N. hemsleyana* (left) and *N. rafflesiana* t.f. (right).

### Pst–Fst comparison in the sister species *N. hemsleyana* and *N. rafflesiana* t.f.

We found that functionally important pitcher shape traits were more diverged than expected under neutral evolution (drift-mutation), including the relative distance between peristome and the "girdle" (the waxy zone resp. tubular section),

as well as the relative internal and external diameters of the pitcher orifice (Figure 5, detailed trait values and confidence intervals for Pst and Fst see Supporting Information S10). Signatures of phenotypic selection were also found for pitcher wall thickness, the relative external orifice diagonal (perpendicular to the other orifice measurement), and for absolute pitcher length. However, critical c/h2 values larger than 0.2 for these traits indicated a lack of robustness: relatively small violations of the assumption that within- and between-species heritabilities are similar could give false-positive results. The relative distance between peristome and digestive fluid level was marginally overlapping with the neutral expectation. One analysed trait, relative internal orifice diagonal, had Pst–Fst $< 0$, but the confidence interval overlapped with zero (Figure 5).

## Discussion

All seven *Nepenthes* morphospecies studied here correspond to distinct genetic clusters in sympatry (Mallet 1995). Among them are the two "forms" of *N. rafflesiana* that are only informally described. The molecular verification of hybrids between at least six species pairs is direct evidence for incomplete reproductive isolation and suggests that individuals with intermediate trap morphology, which were previously observed in the field (Cheek & Jebb 2001; Clarke 2001) may also represent hybrids. However, the identification of hybrids based on trap morphology may not be reliable as several backcrossed hybrids were morphologically not distinguished from parental species in the field. These individuals were only identified as (presumably late generation) hybrids using molecular analysis, illustrating that such hybrids may be more common in the wild than previously anticipated.

### Reticulated evolution

None of the *Nepenthes* species and populations showed monophyly in a genome-wide sense: highly significant alternative relationships (ABBA-BABA tests) were hidden below the bifurcating phylogenetic tree, although this tree received strong support (Figure 3 left). This tree can be interpreted as the majority or average phylogenetic history (Mallet et al. 2016) since it was based on random samples of loci across the genome. However, a substantial minority of loci in the *Nepenthes* genomes has other histories (reticulation) contributed through introgression, beyond what is expected for a bifurcating tree with incomplete lineage sorting.

An important limitation of ABBA-BABA tests is that they do not necessarily identify a specific introgression event: an ABBA-BABA signal between two taxa may instead also reflect allele sharing due to introgression among ancestors or with genetically similar taxa. However, these are not "false positive" introgression signals but they reflect genetic similarity. For example, assume taxon A1 introgressed into B, then ABBA-BABA would be significant not only for A1–B but also for comparisons with a close sister species of A1, e.g. A2–B. Moreover, there might be little evolutionary difference between the two hypotheses that A1 itself or its sister species introgressed into B, because the transferred genes were

most likely highly similar, for otherwise they could not affect ABBA-BABA in both comparisons. This "spillage" effect of introgression on ABBA-BABA tests should decay with genetic distance until no introgressed alleles are shared. In essence, the pairwise f_d matrix reveals pervasive introgression in *Nepenthes* but – at the current stage of the method – cannot pinpoint when and where it happened. More complicated ABBA-BABA-like tests exist to identify introgressing taxa and the direction of gene flow in phylogenetic trees with up to five taxa or populations (Eaton & Ree 2013; Eaton et al. 2015; Pease & Hahn 2015). If more than 4-5 entities were available, these and many other studies also conducted multiple ABBA-BABA tests on preselected combinations and presented test results in tabular form. Our adaptation offers (1) testing of arbitrarily large phylogenetic trees and intuitive visualisation of genetic similarity due to introgression, and (2) an objective discovery of introgression signals, without bias towards preselected hypotheses.

The phylogenetic perspective including ABBA-BABA tests cannot directly reveal timing and duration of introgression. For example, rare gene flow between distant geographic locations, or introgression facilitated by recent anthropogenic habitat alterations could also cause ABBA-BABA signals. In these scenarios, gene flow is too rare and weak to challenge species integrity and divergence. The coalescent models were more informative in this respect.

**Evidence for long-term divergence with gene flow**

Our results from coalescent modelling suggest that gene flow extended over long periods of time and up to the present in many sympatric *Nepenthes* species pairs. In addition to the direct ABC model choice results, several further arguments also shed light on introgression and DGF.

Three species pairs were observed both in Brunei and Singapore but the consistent models and the parameter estimates were different between these locations. For example, data supported DGF for *N. ampullaria–N. gracilis* and *N. ampullaria–N. rafflesiana* in Brunei but not in Singapore. This suggests that (1) the strength of introgression differed between locations, and (2) introgression was stronger or more recent than the spatial genetic isolation between Brunei and Singapore (distance c. 1,200 km; sea-level rise after the last glacial maximum separated Borneo from Singapore <19,000 years ago, Woodruff 2010). Furthermore, species integrity (morphology, genome-wide average "monophyly" in the species tree Figure 3) was preserved despite hybridisation and backcrossing.

DGF is furthermore supported by ancestral population size estimates (Table 2). They were consistently larger than either descendant, indicating lower coalescence rates. Several examples of this effect are known from isolation-with-migration models of speciation (listed in Yang et al. 2017). Hence, ancestral species were unusually diverse in our models, possibly because they represent mixtures of divergent populations or species (Won et al. 2005; Pinho & Hey 2010). Coalescent models strongly suggest that multiple species simultaneously undergo DGF in Brunei. Each species in the Brunei assemblage showed DGF-signatures against

two to six others, forming at least six and up to 15 DGF species pairs. One may argue that the actual number of DGF processes is lower because true donor species can be confused with unobserved closely related species (ghosts). We accounted for this type of confusion in the form of the "ghost" model but the ghost concept can also be interpreted as an indicator of simultaneous gene flow between more than two species: nine species pairs displayed both DGF- and ghost-signatures.

Discrepancies in the pairwise divergence time estimates (Table 2) also hint at multiple-partner introgression and DGF. The estimates ranged from approximately 20k to more than 100k generations but these estimates are only useful for relative comparisons, mainly because mutation rates and generation times in years are unknown for *Nepenthes*. Strikingly, these time estimates cannot be reconciled with the genome-wide concatenation species tree (Figure 3) as the 95% CIs of sister species against third species did frequently not overlap. For example, *N. rafflesiana* t.f.–*N. rafflesiana* g.f. split c. 100,000 generations before *N. hemsleyana* split from *N. rafflesiana* g.f. although the phylogenetic tree grouped *N. hemsleyana*–*N. rafflesiana* t.f. as sisters. Moreover, the split times imply the paradox that the sister species do not first coalesce (going backward in time) with each other but instead first with the third species *N. rafflesiana* g.f. Methodological artefacts are unlikely since these estimates were reasonably precise (cross-validation error c. 5-12%, c. 30% in some pairs with *N. bicalcarata*). Importantly, however, such discrepancies are entirely expected under introgression: the time to interspecific coalescence should vary along genomes (e.g. Mailund et al. 2014), and the minimum coalescent time in the genome should be reduced relative to less- or non-introgressing pairs (as exploited as an introgression test by Meyer et al. 2016). The split time estimates could also be influenced by gene flow from ghost species that occurred in addition to gene flow within a focal pair.

In essence, a comprehensive interpretation of pairwise coalescent modelling results is very difficult but they clearly support the notion of long-term interspecific gene flow among multiple species. We posit that gene flow between sympatric *Nepenthes* species studied here has occurred over long time periods, either in primary divergence or old secondary contacts. This argumentation is based on the fact that all seven species are jointly distributed over large geographic distances (within Borneo, and between Borneo and Singapore). Such distributions require at least several hundred generations of mutual range invasion (sympatry), even under the extreme assumption that species originated in strict isolation at their current range margins.

The maintenance of species divergence in the face of long-term gene flow suggests that RI in sympatric *Nepenthes* is affected by diversifying selection. RI with a selectively neutral basis, such as classical Dobzhansky-Muller incompatibilities, is quickly lost in secondary contact when RI is not complete (Lindtke & Buerkle 2015). Consequently, we propose that these *Nepenthes* co-exist in

migration-selection equilibrium (Lenormand 2002), because sympatry is old and gene flow had ample time to reverse speciation but failed to do so.

**What factors drive divergence in the face of gene flow in *Nepenthes*?**
*Nepenthes* pitcher traps are highly diverse and often differ more strongly between species than other vegetative (e.g. leaf shape) and most reproductive structures, as exemplified by identification keys (Cheek & Jebb 2001). Pitcher morphology is likely to be under diversifying selection between the sister species *N. hemsleyana* and *N. rafflesiana* t.f. because their neutral genetic differentiation is much too low to explain their divergent trap phenotypes (Figure 5). Previous work described presumably adaptive differences in morphological, ecological and physiological pitcher traits (Clarke et al. 2011; Grafe et al. 2011; Schöner et al. 2013, 2015b; a; Lim et al. 2015). In particular, the upper part of the pitchers, which serves to acoustically attract and provide shelter to mutualistic bats in *N. hemsleyana* but not in *N. rafflesiana* t.f. (Schöner et al. 2015b), is highly diverged. The robust Pst–Fst comparison clearly supports diversifying selection on multiple pitcher morphology traits in this species pair, although DGF was not the only plausible speciation history, presumably due to very low genetic divergence.

In the following we introduce a hypothesis that may link RI with diversifying selection and the *Nepenthes* pitcher traps. It has previously been speculated that competition for prey could drive divergence in trap functions and prey spectra (Moran et al. 1999; Bauer et al. 2012; Chin et al. 2014; Gaume et al. 2016). Indeed, reliance on a shared limiting resource promotes resource partitioning (character displacement) and adaptive speciation through frequency-dependent fitness of phenotypes (Dieckmann & Doebeli 1999; Doebeli & Dieckmann 2000; Schluter 2000). However, in a typical undisturbed habitat, *Nepenthes* density is so low that they almost never directly interact above or below ground. Furthermore, we believe that the abundance of prey (arthropods, leaf litter) is too high for pitcher plants to ever have to compete over them (however, this might be different in high elevation habitats, and for mutualistic interactions with mammals). Further, for sympatric assemblages of other carnivorous plants (*Sarracenia*, *Drosera*), available evidence suggests absence of competition for prey (Ellison & Gotelli 2009). We argue that *Nepenthes* are effectively not density-limited by congeners, and that competition for prey cannot explain adaptive diversification in pitcher traps. Instead, we propose that competition for light from surrounding vegetation and the energetic balance of carnivory should determine *Nepenthes* fitness, in line with the cost-benefit model for the evolution of plant carnivory (Givnish et al. 1984; Ellison 2006; Pavlovič & Saganová 2015). This model posits that carnivorous plants are able to compete against other plants when the photosynthetic gain from prey-derived mineral supply exceeds the energetic costs of carnivory. We put forward that different types of traps achieve similar net-benefits, alluding to a "rugged" fitness landscape with multiple optima (Gavrilets 1997; Schluter 2000). Thus, sympatric *Nepenthes* species may represent alternative solutions to multi-dimensional cost-benefit trade-offs, each allowing existence in the same

"harsh environment" through mutually exclusive combinations of functional morphology and physiology. The intermediate trap phenotypes of hybrids may be maladaptive, establishing post-zygotic extrinsic RI. Multi-dimensional trade-offs can indeed drive speciation, and thus also maintain divergence in the face of gene flow, if linkage disequilibrium establishes between selected quantitative trait loci (QTLs) and assortative mating QTLs (Artzy-Randrup & Kondrashov 2006). Finally, our hypothesis interprets *Nepenthes* prey spectra as by-products of economically viable trait combinations, in contrast to all previous work on the topic which invariably considered prey spectra to indicate resource partitioning (Moran et al. 1999; Bauer et al. 2011; Chin et al. 2014; Gaume et al. 2016).

As an experimental test of our hypothesis, we expect that *Nepenthes* hybrids have little reproductive success due to a poor energy balance of carnivory in intact habitats (post-zygotic extrinsic RI). Hybrid fitness should increase the less important carnivory, which could be the case in disturbed habitats with increased light and soil nutrient availability and reduced competition. Early generation hybrids are indeed more frequently observed in logged areas, roadsides and landslides (pers. obs.; Peng & Clarke 2015), and in some cases capture fewer insects than either parental species (Peng & Clarke 2015). Natural introgression, despite apparent selection against hybrids, may then be explained by transient hybrid fitness during periods of localised natural habitat disturbance, such as fires and landslides. This would constitute a diagnostic test of ecological speciation, in which the strength of postzygotic RI depends on environmental conditions (Schluter 2000). Reciprocal transplant experiments of parental species and hybrids in different habitats – primary vegetation versus disturbed sites – could be performed to test this hypothesis (Favre et al. 2017).

Alternative hypotheses that do not involve diversifying selection on pitchers must explain pitcher divergence in the face of gene flow and the occurrence of (partial) RI by other mechanisms. Other traits may be under diversifying selection and pitcher divergence a by-product of shared genetic bases (pleiotropy). For example, diversifying selection and extrinsic RI can result from habitat differences, which usually underlie ecological speciation (Nosil 2012). However, neither spatial segregation nor physical differences between the stands of five of seven *Nepenthes* in Brunei are obvious. Only two species might prefer different habitats: *N. rafflesiana* g.f. was found exclusively in the drier heath forests, and *N. mirabilis* tolerates inundation at river margins. Pitcher QTLs might also overlap with assortative mating QTLs (e.g. pollinator-driven speciation, phenology), or with loci conferring intrinsic mechanisms of RI, such as hybrid breakdown (e.g. Renaut & Bernatchez 2011). However, no evidence in support of such scenarios has thus far been reported.

As ongoing DGF appears likely among multiple sympatric *Nepenthes* species that span a wide range of divergence times and gene flow intensities, the genus emerges as a promising system to disentangle the relative importance of neutral and adaptive forces during introgression and speciation (Cruickshank & Hahn

2014; Payseur & Rieseberg 2016). We demonstrate that *Nepenthes* species boundaries are "semi-permeable" (Harrison & Larson 2014) and that divergence was frequently challenged by gene flow. This provides the foundation for further investigation of additional traits and genomic regions that resist homogenisation, the selective forces acting upon them, and whether loci underlying the carnivorous syndrome are also be causative of postzygotic extrinsic RI and thus of ecological speciation.

**General conclusions**

Introgression in species complexes can help unravel adaptive diversification, which may be common in nature. Yet adaptive diversification in the presence of gene flow has been studied in detail for only few organisms, perhaps because methods to detect these processes are few. We adapted a phylogenomic (D-test) and a population genomic approach (coalescent and ABC framework) to test for ancient and ongoing gene flow among multiple species using genome-wide markers. To our knowledge, this is the first study to address DGF explicitly in any plant group inhabiting megadiverse tropical forests, where species-rich communities of close relatives are common.

## Acknowledgements

We thank Abdul Hadzid Harith Tinggal, Ieney Daud (Telamba Homestay) and their family for hospitality in Brunei, and Lam Weng Ngai for assistance with field work in Singapore. Hirzi Luqman and Christoph Küffer are thanked for sampling in the Seychelles. The Ministry of Industry and Primary Resources of Brunei Darussalam, the Agri-Food and Veterinary Authority and the National Parks Board of Singapore, and the Seychelles Agricultural Agency kindly granted permissions to conduct research and export plant material. We thank Claudia Michel for excellent wet lab support. The Genetic Diversity Center Zürich, the Quantitative Genomics Facility Basel, and the Functional Genomics Center Zürich are thanked for sequencing and computers. Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies. J. Stapley and K. Csilléry provided helpful comments on the manuscript. This work was funded by the ETH Zürich and partly supported through SNF grant no 31003A_160123 to AW.

## References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.

Artzy-Randrup Y, Kondrashov AS (2006) Sympatric speciation under incompatibility selection. *Proceedings of the National Academy of Sciences*, **103**, 11619-11624.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113-148.

Bauer U, Clemente CJ, Renner T, Federle W (2012) Form follows function: Morphological diversification and alternative trapping strategies in carnivorous *Nepenthes* pitcher plants. *Journal of Evolutionary Biology*, **25**, 90-102.

Bauer U, Grafe TU, Federle W (2011) Evidence for alternative trapping strategies in two forms of the pitcher plant, *Nepenthes rafflesiana*. *Journal of Experimental Botany*, **62**, 3683-3692.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in population genetics. *Genetics*, **162**, 2025-2035.

Beerli P (2004) Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, **13**, 827-836.

Brommer JE (2011) Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. *Journal of Evolutionary Biology*, **24**, 1160-1168.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684-687.

Camacho-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA (2013) Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, **13**, 663-673.

Cheek MR, Jebb MHP (2001) Nepenthaceae. *Flora Malesiana*, **15**, 1-157.

Cheek M, Jebb M (2013) The *Nepenthes micramphora* (Nepenthaceae) group, with two new species from Mindanao, Philippines. *Phytotaxa*, **151**, 25-34.

Chin L, Chung A, Clarke C (2014) Interspecific variation in prey capture behavior by co-occurring *Nepenthes* pitcher plants: Evidence for resource partitioning or sampling-scheme artifacts? *Plant Signaling & Behavior*, **9**, e27930.

Christe C, Stölting KN, Bresadola L et al. (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology*, **25**, 2482-2498.

Christe C, Stölting KN, Paris M et al. (2017) Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, **26**, 59-76.

Clarke CM (1992) *The ecology of metazoan communities in Nepenthes pitcher plants in Borneo, with special reference to the community of Nepenthes bicalcarata Hook. f.* PhD Thesis. University of New England, Armidale, Australia.

Clarke CM (1997) *Nepenthes of Borneo*. Natural History Publications (Borneo), Kota Kinabalu.

Clarke CM (2001) *Nepenthes of Sumatra and Peninsular Malaysia*. Natural History Publications (Borneo), Kota Kinabalu.

Clarke C, Moran JA, Lee CC (2011) *Nepenthes baramensis* (Nepenthaceae) - a new species from north-western Borneo. *Blumea*, **56**, 229-233.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-3157.

Csilléry K, François O, Blum MGB (2012) abc: An R package for Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution*, **3**, 475-479.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, **24**, 332-340.

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354-357.

Doebeli M, Dieckmann U (2000) Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *The American Naturalist*, **156**, S77-S101.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239-2252.

Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**, 2587-2601.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, **62**, 689-706.

Ellison AM (2006) Nutrient limitation and stoichiometry of carnivorous plants. *Plant Biology*, **8**, 740-747.

Ellison AM, Gotelli NJ (2009) Energetics and the evolution of carnivorous plants – Darwin's 'most wonderful plants in the world.' *Journal of Experimental Botany*, **60**, 19-42.

Favre A, Widmer A, Karrenberg S (2017) Differential adaptation drives ecological speciation in campions (*Silene*): Evidence from a multi?site transplant experiment. *New Phytologist*, **213**, 1487-1499.

Fontaine MC, Pease JB, Steele A et al. (2015) Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, **347**, 1258524.

Gaume L, Bazile V, Huguin M, Bonhomme V (2016) Different pitcher shapes and trapping syndromes explain resource partitioning in *Nepenthes* species. *Ecology and Evolution*, **6**, 1378-1392.

Gavrilets S (1997) Evolution and speciation on holey adaptive landscapes. *Trends in Ecology & Evolution*, **12**, 307-312.

Givnish TJ, Burkhardt EL, Happel RE, Weintraub JD (1984) Carnivory in the bromeliad *Brocchinia reducta*, with a cost/benefit model for the general restriction of carnivorous plants to sunny, moist, nutrient-poor habitats. *The American Naturalist*, **124**, 479-497.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111-2127.

Grafe TU, Schöner CR, Kerth G, Junaidi A, Schöner MG (2011) A novel resource-service mutualism between bats and pitcher plants. *Biology Letters*, **7**, 436-439.

Green RE, Krause J, Briggs AW et al. (2010) A draft sequence of the Neandertal genome. *Science*, **328**, 710-722.

Gross BL, Rieseberg LH (2005) The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, **96**, 241-252.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, **5**, e1000695.

Harrison RG (2012) The language of speciation. *Evolution*, **66**, 3643-3657.

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, **105**, 795-809.

Juniper BE, Robins RJ, Joel DM (1989) *The carnivorous plants*. Academic Press, London.

Lamichhaney S, Berglund J, Almén MS et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371-375.

Leaché AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A (2015) Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology*, **64**, 1032-1047.

Leinonen T, McCairns RJS, O'Hara RB, Merilä J (2013) QST-FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, **14**, 179-190.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183-189.

Lim YS, Schöner CR, Schöner MG et al. (2015) How a pitcher plant facilitates roosting of mutualistic woolly bats. *Evolutionary Ecology Research*, **16**, 581-591.

Lindtke D, Buerkle CA (2015) The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, **69**, 1987-2004.

Lindtke D, Gompert Z, Lexer C, Buerkle CA (2014) Unexpected ancestry of *Populus* seedlings from a hybrid zone implies a large role for postzygotic selection in the maintenance of species. *Molecular Ecology*, **23**, 4316-4330.

Mailund T, Munch K, Schierup MH (2014) Lineage sorting in apes. *Annual Review of Genetics*, **48**, 519-535.

Mallet J (1995) A species definition for the modern synthesis. *Trends in Ecology & Evolution*, **10**, 294-299.

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays*, **38**, 140-149.

Martin SH, Dasmahapatra KK, Nadeau NJ et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817-1828.

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**, 244-257.

Mayr E (1942) *Systematics and the origin of species from the view-point of a zoologist*. Columbia University Press, New York.

Meier JI, Marques DA, Mwaiko S et al. (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, **8**, 14363.

Meimberg H, Heubl G (2006) Introduction of a nuclear marker for phylogenetic analysis of Nepenthaceae. *Plant Biology*, **8**, 831-840.

Meimberg H, Wistuba A, Dittrich P, Heubl G (2001) Molecular phylogeny of Nepenthaceae based on cladistic analysis of plastid trnK intron sequence data. *Plant Biology*, **3**, 164-175.

Meyer BS, Matschiner M, Salzburger W (2016) Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic Biology*, syw069.

Moran JA, Booth WE, Charles JK (1999) Aspects of pitcher morphology and spectral characteristics of six Bornean *Nepenthes* pitcher plant species: Implications for prey capture. *Annals of Botany*, **83**, 521-528.

Moran JA, Clarke CM (2010) The carnivorous syndrome in Nepenthes pitcher plants. *Plant Signaling & Behavior*, **5**, 644-648.

Mullins JT (2000) *Molecular Systematics of Nepenthaceae*. PhD Thesis. University of Reading, Reading, UK.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269-5273.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford; New York.

Pavlovič A (2012) Adaptive radiation with regard to nutrient sequestration strategies in the carnivorous plants of the genus *Nepenthes*. *Plant Signaling & Behavior*, **7**, 295-297.

Pavlovič A, Saganová M (2015) A novel insight into the cost-benefit model for the evolution of botanical carnivory. *Annals of Botany*, **115**, 1075.

Payseur BA, Rieseberg LH (2016) A genomic perspective on hybridization and speciation. *Molecular Ecology*, **25**, 2337-2360.

Pease JB, Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, **64**, 651-662.

Peng HS, Clarke C (2015) Prey capture patterns in *Nepenthes* species and natural hybrids - are the pitchers of hybrids as effective at trapping prey as those of their parents? *Carnivorous Plant Newsletter*, **44**, 62-79.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE*, **7**, e37135.

Pinho C, Hey J (2010) Divergence with gene flow: Models and data. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 215-230.

Price AL, Patterson NJ, Plenge RM et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904-909.

Pudlo P, Marin J-M, Estoup A et al. (2016) Reliable ABC model choice via random forests. *Bioinformatics*, **32**, 859-866.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e431.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, **16**, 359-371.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational inference of population structure in large SNP datasets. *Genetics*, **197**, 573-589.

Renaut S, Bernatchez L (2011) Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity*, **106**, 1003-1011.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380-390.

Roux C, Fraisse C, Castric V et al. (2014) Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, **27**, 1662-1675.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574-1587.

Scharmann M, Grafe TU (2013) Reinstatement of *Nepenthes hemsleyana* (Nepenthaceae), an endemic pitcher plant from Borneo, with a discussion of associated *Nepenthes* taxa. *Blumea*, **58**, 8-12.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, New York, USA.

Schöner CR, Schöner MG, Kerth G, Grafe TU (2013) Supply determines demand: Influence of partner quality and quantity on the interactions between bats and pitcher plants. *Oecologia*, **173**, 191-202.

Schöner CR, Schöner MG, Kerth G, Suhaini SN binti P, Grafe TU (2015a) Low costs reinforce the mutualism between bats and pitcher plants. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, **258**, 1-5.

Schöner MG, Schöner CR, Simon R et al. (2015b) Bats are acoustically attracted to mutualistic carnivorous plants. *Current Biology*, **25**, 1911-1916.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198-207.

Simmons MP, Norton AP (2014) Divergent maximum-likelihood-branch-support values for polytomies. *Molecular Phylogenetics and Evolution*, **73**, 87-96.

Simonsen M, Pedersen CNS (2011) Rapid computation of distance estimators from nucleotide and amino acid alignments. In: *Proceedings of the 2011*

*ACM Symposium on Applied Computing SAC '11.*, pp. 89-93. ACM, New York, NY, USA.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology*, **14**, 67-73.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312-1313.

Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505-518.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94-98.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLOS Biology*, **3**, e285.

Vonlanthen P, Bittner D, Hudson AG et al. (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357-362.

Wakeley J (2009) *Coalescent theory: an introduction*. Roberts & Company Publishers, Greenwood Village, Colorado, USA.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Won Y-J, Sivasundar A, Wang Y, Hey J (2005) On the origin of Lake Malawi cichlid species: A population genetic analysis of divergence. *Proceedings of the National Academy of Sciences*, **102**, 6581-6586.

Woodruff DS (2010) Biogeography and conservation in Southeast Asia: How 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodiversity and Conservation*, **19**, 919-941.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851-865.

Yang M, He Z, Shi S, Wu CI (2017) Can genomic data alone tell us whether speciation happened with gene flow? *Molecular Ecology*, **26**, 2845-2849.

## Author Contributions

MS carried out field work, bioinformatics and statistical analyses, and some of the molecular lab work; FM and TUG contributed logistic support and materials; MS and AW conceived and designed the study, and wrote the manuscript.

# Chapter 1 Supporting Information S1, S2, S3, S4, S7

further Supporting Information is available on the CD attached to this thesis

## S1: ddRAD-seq bioinformatics details

We processed the raw sequences with a modified dDocent pipeline (Puritz et al. 2014). Sequencing reads were filtered for any sequences used in library prep (adapters etc) with trimmomatic (Bolger et al. 2014), and discarded when any of the first five bases (the ddRAD-barcode) had Phred-scores $< 20$ (custom python script). Only after this step we demultiplexed and quality-filtered using process_radtags of the Stacks pipeline (Catchen et al. 2013, "process_radtags –inline_index -E phred33 -e ecoRI -c -q -w 0.15 -s 20"): reads were discarded when containing any ambiguous bases, less than mean Q20 (error probability 1%) in a window of 15% read length, or when the restriction site was absent.

Since we had single-end reads for most species, we modified the dDocent pipeline to de novo assemble reference contigs based on single-end reads. All reads of all individuals were concatenated and the read depth for each unique sequence was counted. Unique sequences with at least 10 identical reads were retained, and clustered using vsearch (https://github.com/torognes/vsearch) v1.1.1 at 90% identity. Clusters containing more than 100 unique sequences (indicative of repetitive content / high level paralogy) were discarded. Each remaining cluster was globally re-aligned using MUSCLE (Edgar 2004) v3.8.31, and contigs exported as the majority consensus sequence. Consequently, we removed potentially redundant contigs by collapsing them at 90% identity using Cd-hit v4.6 (Li & Godzik 2006). We refer to the surviving reference contigs as RAD-tags.

Reads were mapped against this reference using BWA-MEM (Li 2013, bwa mem -L 20,5 -M -T 10 -A 1 -B 4 -O 6), and filtered for minimum mapping quality of 1 using samtools (Li et al. 2009), hence removing ambiguously mapping reads. Population variants were called using FreeBayes (Garrison & Marth 2012) v1.0.2-6-g3ce827d-dirty (freebayes -m 5 -q 5 -E 3 -G 3 –min-repeat-entropy 1 -V –hwe-priors-off -i -X -u). This step was run in parallel threads on subsets of the RAD-tags, which drastically reduced the runtime.

Raw variant calls in .vcf format were quality-filtered using vcftools (Danecek et al. 2011) and custom scripts, as designed by J. Puritz and by ourselves (genotypes: minimum coverage = 3 and maximum coverage = mean individual coverage + 2SD of mean individual coverage; sites: allele balance $> 0.25$ and $< 0.75$ or $< 0.01$; quality/depth ratio $> 0.25$; genotyping quality $\geq$ mean depth plus three times the square of the mean depth; removal of all RAD-tags with excess heterozygosity at $p \leq 0.05$ in the Hardy-Weinberg test). Only biallellic SNPs were retained.

Further filtering on per-site missingness across individuals and/or populations was specifically applied for different downstream analyses. Individuals with less than c. 52,000 genotyped SNPs (> 96% missingness) were excluded.

## S2: Details on multiple ABBA-BABA tests and pairwise f_d matrix

Genotypes were converted to derived allele frequencies polarised by N. pervillei. Sites were filtered for presence in at least 50% of samples per population, while for the individual-based tests all sites present in each of the four samples were used, i.e. each test was based on a different set of sites. Tests were implemented in a custom script based on Dtest.py of PyRAD (Eaton 2014), while D and f_d were calculated according to equations (1-3) resp. (6) from Martin et al. (2015).

We built D and f_d statistics over all available sites, thus "genome-wide" and not for small genomic regions or individual loci, which can be problematic (Martin et al. 2015). Like PyRAD (Eaton 2014), we assessed the significance of the deviation of the D-statistic from zero by approximating the standard deviation through 1,000 bootstrap replicates from the available sites, calculating a Z-score, and transformation to a two-sided p-value. Bootstrapping estimates much larger standard deviations than the block jackknife-estimate used in other studies and is thus more conservative.

Two significance thresholds were enforced on ABBA-BABA test results: a minimal number of 500 ABBA-BABA informative sites, and a maximum p-value for the D-statistic (0.05 for the population level analysis and 0.0001 for the individual-level analysis, setting f_d=0 if not met). We then summarised all significant D-tests conducted on a backbone tree as a pairwise matrix of f_d between all populations/individuals (see main text for "pairwise" interpretation of the D-test). If multiple tests were significant for a given pair, we retained only the maximum observed f_d. Pairwise f_d matrices were plotted in R.

Pairwise f_d matrices are conceptually similar to the pairwise genetic covariance residuals from TreeMix (Pickrell & Pritchard 2012) which are also plotted the same way. We explored TreeMix but found that its power to detect migration was low in our dataset. This can be attributed to TreeMix's requirement that each SNP be genotyped in all populations, such that for RAD-seq data only few SNPs remain (moreover, these are biased towards low divergence). Between-SNP heterogeneity in divergence, the common underlying principle of migration detection in both TreeMix and ABBA-BABA, is more difficult to detect with fewer available SNPs. In contrast to a global TreeMix analysis, multiple ABBA-BABA tests allow to use different sets of SNPs in each test. The advantage is that we can harness the hundreds of thousands of SNPs passing filtering criteria in at least four populations, rather than restricting us to the few thousand SNPs passing the missingness filters for all 11 populations simultaneously. A possible criticism to our approach is that different sets of loci are tested in each ABBA-BABA test and hence the tests might not be comparable –however, any large sets of loci sampled at random from the same genomes are expected to contain similar information,

by the law of large numbers. This argument also justifies the application of multiple pairwise coalescent models rather than one joint single model with all populations.

Other model-free methods to infer introgression such as the f4 statistic (Reich et al. 2009; Meyer et al. 2016) are essentially per-site statistics which do not take genomic heterogeneity into account –each SNP on its own must show an introgression signal to influence the genome-wide average f4, and the statistical significance. It works very well when population divergence is low and when the polymorphisms segregate in all of them. But f4 has no power to detect introgression for sites that are not polymorphic on both sides of the root of the assumed 4-taxon tree. Thus, relevant cases of introgression are missed, i.e. immigrant alleles that became fixed in the recipient population, and introgression between populations or species that diverged (drifted) such that polymorphisms are now mostly fixed.

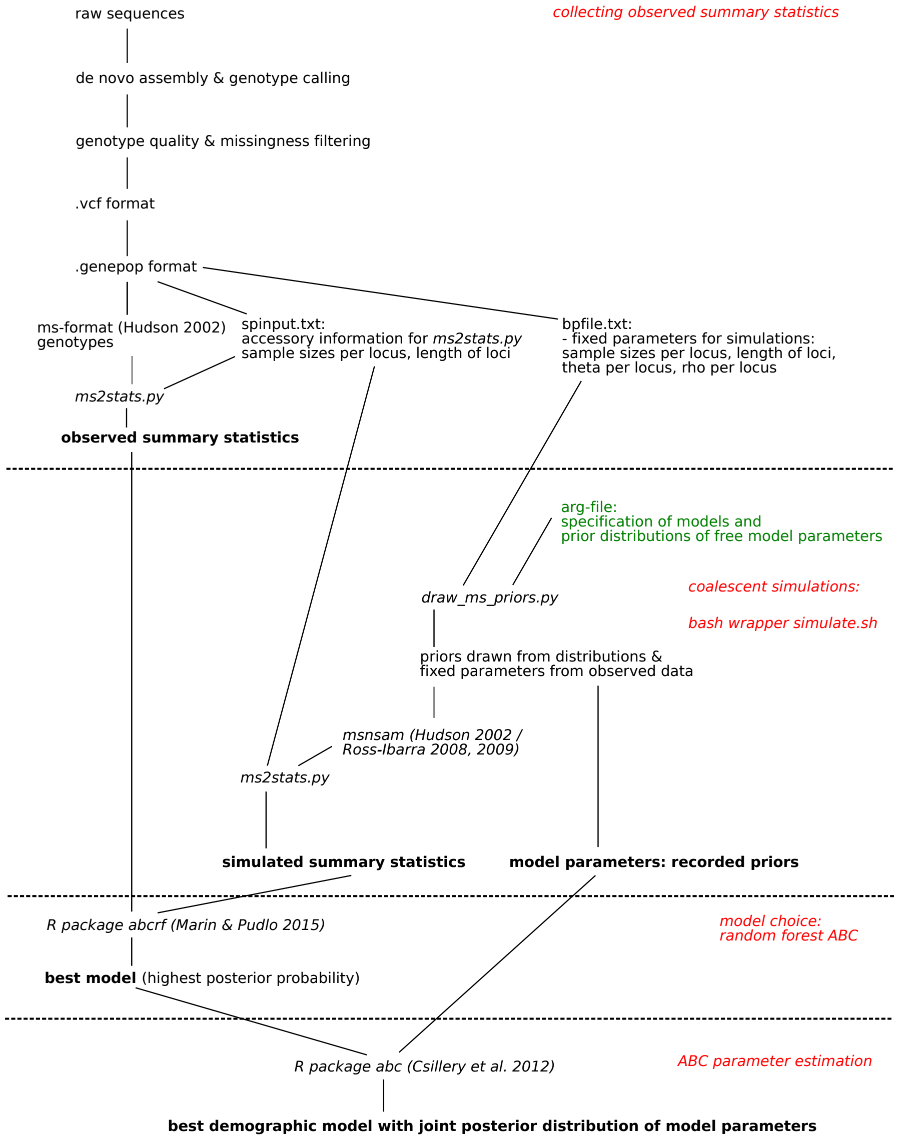
## S3: Details on coalescent modelling and Approximate Bayesian Computation

The general design of our pipeline closely follows the approach pioneered by C. Roux (Roux et al. 2013, http://www.abcgwh.sitew.ch), but was amended and re-written in Python, bash and R (Supporting Information Figure S3-1). Ghost species were implemented as ordinary populations in the coalescent simulations, with the only difference that no sequences were sampled from them.

### Model parameters and priors

Fixed (i.e. not estimated) simulation parameters were the sampling scheme as directly extracted from filtered .vcf genotypes of the observed data (number of RAD-tags, sample size per population per RAD-tag, length in nucleotides per RAD-tag), and the spontaneous population mutation rate theta = $4 \times N \times \mu$= $4 \times 100{,}000 \times 2.5 \times 10^{-8}$ per site per generation. Recombination rates were set to zero, as the RAD-tags were only c. 100 nucleotides long. Note that the ambiguous values for reference population size N = 100,000 and per site per generation mutation rate μ have no effects on our conclusions, because we compared relative population genetic patterns. Our parameters cannot confidently be converted to biologically meaningful absolute units (e.g. years, numbers of plants).

Free model parameters and their prior ranges are listed in Supporting Information Table S3-1. All priors were uniformly distributed, except for some of the time-priors that were necessarily truncated by the precedence of other time-priors (e.g. migration cannot occur before speciation). Suitable priors for effective population sizes and split times were found by preliminary parameter estimation trials, and adjusted so that posteriors would fall comfortably within the bounds. Other approaches to derive "Ne" were not applicable because in our models it is a nuisance parameter (to capture temporal fluctuations and population substructure in space, age and gender) with no predictable relationship to population census. The speciation time prior included the recent past (100 generations ago)




**Supporting Information Figure S3-1.** Flowchart showing all steps, important intermediate files and tools used in the coalescent modelling and ABC pipeline

**Supporting Information Table S3-1.** Prior ranges in coalescent units for the three speciation models (tool ms, Hudson 2002; reference Ne = 100,000). The ghost model included an additional "prior" to specify the ghost topology, i.e. the external and two internal ghosts were sampled at equal probabilities each.

| | DGF | isolation | ghost |
|---|---|---|---|
| *Ne parameters:* | | | |
| N1 | 0.01 – 2 | 0.01 – 2 | 0.01 – 2 |
| N2 | 0.01 – 2 | 0.01 – 2 | 0.01 – 2 |
| Nghost | - | - | 0.01 – 2 |
| Nx | - | - | 0.01 – 2 |
| Nmrca | 0.01 – 2 | 0.01 – 2 | 0.01 – 2 |
| *time parameters:* | | | |
| T_merge_2_into_1 | 0.00025 – 2 | 0.00025 – 2 | 0.00025 – 2 |
| T_merge_1_into_ghost | - | - | 0.00025 – 2* |
| *migration parameters:* | | | |
| m12_scale | 0 – 20 | - | - |
| m21_scale | 0 – 20 | - | - |
| m1ghost_scale | - | - | 0 – 20 |
| m2ghost_scale | - | - | 0 – 20 |
| m12_prop_mig | 0.05 – 0.95 | - | - |
| m21_prop_mig | 0.05 – 0.95 | - | - |
| m1ghost_prop_mig | - | - | 0.05 – 0.95 |
| m2ghost_prop_mig | - | - | 0.05 – 0.95 |

*condition: T_merge_1_into_ghost > T_merge_2_into_1

up to 800k generations ago, stretching well beyond the most basal split among extant *Nepenthes* which occurred c. 400k generations ago given these priors (parameter estimates for isolation model of *N. pervillei* vs. Asian species, data not shown). Migration rates ("m_scale") covered zero to near-panmixia (panmixia was determined as the migration rate above which netDiv and Fst remained at a near-zero plateau; data not shown). We implemented heterogeneity in migration rates among loci – a conceptual implication of introgression – as a simple binomial probability, given by the parameter "prop_mig", i.e. the proportion of loci having a migration rate greater than zero and of magnitude "m_scale".

In practice, simulation parameters were drawn by a custom python script (draw_ms_priors.py) from the prior distributions (Supporting Information Table S3-1) and supplied to the coalescent simulator msnsam (Hudson 2002; Ross-Ibarra et al. 2008; Roux et al. 2011).

In each of the 24 pairwise comparisons, we simulated c. 40k sets of summary statistics for each of the three models "DGF", "isolation" and "ghost", each one a summary over thousands of RAD-tags according to the sampling schemes of observed data, totalling more than 18.2 billion coalescent simulations.

**Summary statistics**

We selected or developed a total of 114 summary statistics. In general, statistics were averaged over all sites in a RAD-tag. Statistics were then further summarised

over all available RAD-tags (i.e. "genome-wide"). We calculated averages and standard deviations for numbers of segregating sites, numbers of private alleles, numbers of reciprocally fixed sites, nucleotide diversity (Nei & Li 1979), Watterson's theta, Tajima's D, absolute divergence ("dxy", Nei & Li 1979), net divergence ("netdiv", Nei & Li 1979), and Fst (Weir & Cockerham 1984). The similarity between nucleotide diversities of the two species was represented by Pearson's correlation coefficient. To allow a more complete picture of heterogeneity among RAD-tags, we calculated the 5%, 25%, 50%, 75% and 95% percentiles of the dxy, netdiv and Fst distributions. In addition, we constructed 30-bin normalised histograms for each of dxy, netdiv and Fst, and extracted from these the mean bin content of (1) the three lowest bins, (2) the three sub-highest bins, and (3) the three highest bins. These three histogram chunks should contain rough information on the uni- or bimodality of the distributions. For the Fst-distribution, we also included the two shape parameters of a beta-distribution as fitted by the function scipy.stats.beta.fit in python (for this purpose negative Fst estimates were set to zero instead). We constructed the folded 2D or joint site frequency spectrum, also called allele frequency spectrum (Gutenkunst et al. 2009). The normalised content of each of the 55 bins in a 10 x 10 truncated (since folded) two-dimensional histogram of minor allele frequencies (globally defined MAF: range 0.0 to 1.0) was included as summary statistics.

**ABC random forest model choice**

Prior to any ABC inference, we assured that all observed summary statistics could indeed be produced within our model- and parameter space (priors) by inspecting histograms and PCA axes (R function prcomp) of simulated summary statistics against the observed summary statistics.

We used a novel approach to model choice in ABC that partitions the problem into a classification step and a subsequent step for the estimation of a posterior probability, both solved by the Random Forest (RF) machine learning algorithm (Pudlo et al. 2016). RF is superior to earlier implementations of ABC model choice as it requires much fewer simulations and naturally handles complex interactions, collinearity, and non-informative summary statistics. Model choice in abcrf was performed first globally on all three models, and subsequently pairwise against each inferior model. Classification and regression RFs contained 1,000 trees, and used c. 120k (global) resp. c. 80k (pairwise) bootstrap replicates. For each species pair, we present the globally best model, and inferior models if Bayes Factor $<$ 3.0 in pairwise comparisons. If subsequent pairwise comparisons contradicted a global comparison, the offending models were also deemed consistent with observations.

**Parameter estimation**

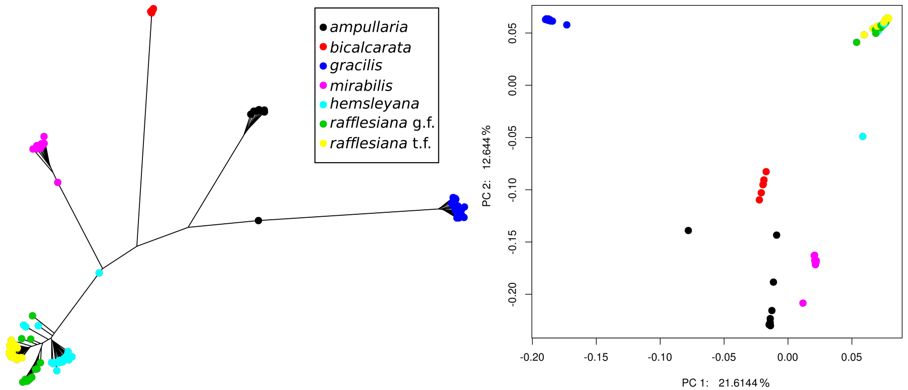
For each species pair, we estimated demographic model parameters for "DGF" if found consistent in model choice, and otherwise for "isolation". "Ghost" was never fitted because the property of interest, the presence or absence of introgres-

sion between sampled species, is represented by the two other models. Parameter space was explored over a total of approximately 40k simulations, and summary statistics with variance >$10^{-7}$ were retained. Using the R package abc (Csilléry et al. 2012), we obtained initial joint posterior parameter distributions by rejection, using a tolerance rate of 1% (i.e. the 400 simulations closest (euclidean distance) to the observed data). Joint posteriors were then adjusted by a multivariate machine learning regression, i.e. 50 neural networks of size 10. All parameters were log-transformed. The cross-validation error was determined in 50 iterations with the same settings.

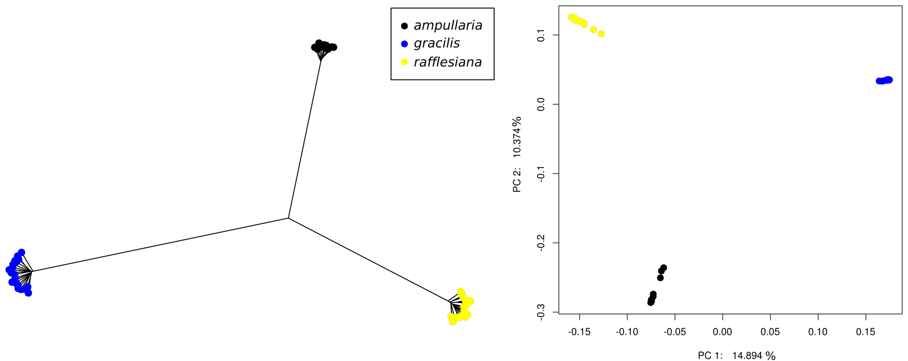
Extensive tests for cross-validation error on our datasets revealed that neural network adjustment based on all summary statistics (up to 114 predictors) outperformed local linear regression based on smaller "optimised" subsets of summary statistics, chosen on highest r-squared in individual statistic - parameter linear regressions, and reduced collinearity (Blum et al. 2013; Shafer et al. 2015). Tolerance rates higher or lower than 1% increased cross-validation error in our datasets.

We simplified the migration parameters in the DGF models by forming the product of the binary and the scale parameters of migration ("prop_mig" × "m_scale"). This combined parameter has a straightforward biological interpretation as the average migration rate across the genome, and in trials exhibited lower cross-validation error than the separate parameters.

## S4: Population genetic structure per location

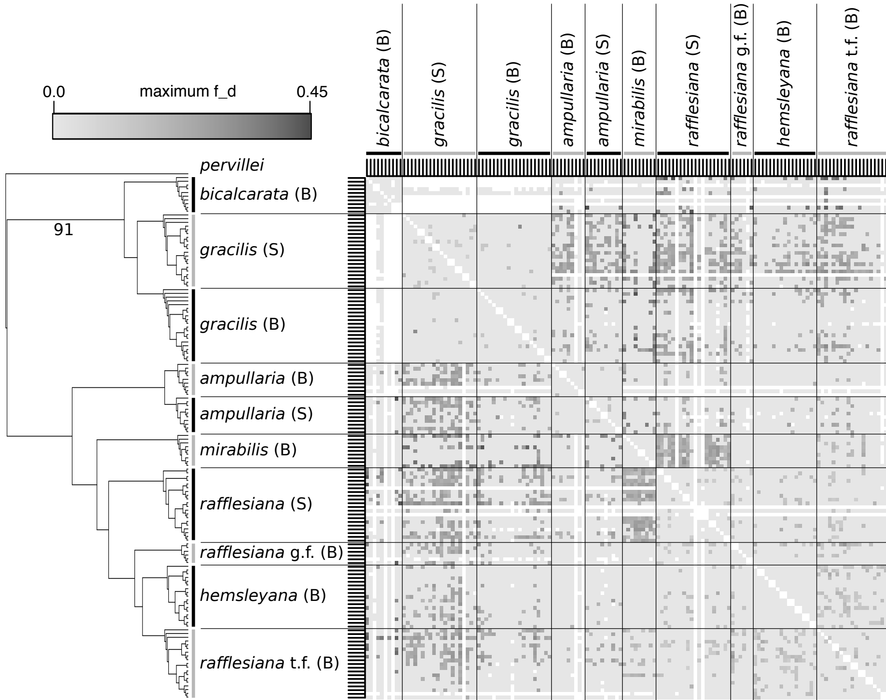



**Supporting Information Figure S4-1.** Brunei *Nepenthes* community: Neighbour Joining tree (left) and first two axes of a Principal Component Analysis (right) for genotypes of seven a-priori determined morphospecies (colour coded). See main text for the distinction between *N. rafflesiana* t.f. and *N. rafflesiana* g.f.




**Supporting Information Figure S4-2.** Singapore *Nepenthes* community: Neighbour Joining tree (left) and first two axes of a Principal Component Analysis (right) for genotypes of three a-priori determined morphospecies (colour coded).

## S7: individual-based phylogenetic tree and f_d matrix




**Supporting Information Figure S7-1.** Phylogenetic topology of *Nepenthes* individuals (left, maximum-likelihood tree, branch lengths not shown) and pairwise matrix showing introgression against this tree (right, maximum f_d statistics from multiple ABBA-BABA tests). All branches except one received full SH-like aLRT support (annotated; not considering branches within populations). Each f_d value greater 0.0 (lightest grey) resulted from a significant ABBA-BABA test ($p \leq 0.0001$) based on at least 500 ABBA-BABA informative sites. Blank cells indicate impossible pairwise tests, or tests excluded due to too few informative sites. Individuals are grouped into their populations (B = Brunei, S = Singapore), indicated by black lines on the pairwise matrix.

## References

Blum MGB, Nunes MA, Prangle D, Sisson SA (2013) A comparative review of dimension reduction methods in Approximate Bayesian Computation. *Statistical Science*, **28**, 189-208.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124-3140.

Csilléry K, François O, Blum MGB (2012) abc: An R package for Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution*, **3**, 475-479.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Eaton DAR (2014) PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844-1849.

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.3907 [q-bio].

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, **5**, e1000695.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337-338.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997 [q-bio].

Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, **32**, 244-257.

Meyer BS, Matschiner M, Salzburger W (2016) Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. *Systematic Biology*, syw069.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269-5273.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, **8**, e1002967.

Pudlo P, Marin J-M, Estoup A et al. (2016) Reliable ABC model choice via random forests. *Bioinformatics*, **32**, 859-866.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e431.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature*, **461**, 489-494.

Ross-Ibarra J, Wright SI, Foxe JP et al. (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLOS ONE*, **3**, e2411.

Roux C, Castric V, Pauwels M et al. (2011) Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLOS ONE*, **6**, e26872.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574-1587.

Shafer ABA, Gattepaille LM, Stewart REA, Wolf JBW (2015) Demographic inferences using short-read genomic data in an approximate Bayesian computation framework: In silico evaluation of power, biases and proof of concept in Atlantic walrus. *Molecular Ecology*, **24**, 328-345.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

# Chapter 2
# Sex-determination and sex chromosomes are shared across the radiation of dioecious *Nepenthes* pitcher plants

---

a manuscript under review co-authored by

M Scharmann[1], T U Grafe[2], F Metali[2], A Widmer[1]

[1] ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology (IBZ), Universitätstrasse 16, 8092 Zurich, Switzerland
[2] Faculty of Science, Universiti Brunei Darussalam, Jalan Tungku Link, Bandar Seri Begawan, BE1410 Brunei Darussalam

## Abstract

Plants with separate sexes (dioecy) represent a minority but dioecy has evolved multiple times independently in plants. Our understanding of sex determination systems in plants and of the ecological factors and molecular changes associated with the evolution of dioecy, however, remain limited. Here, we study the sex-determination system in recalcitrant plants with uniform chromosomes and not amenable to controlled breeding: *Nepenthes* pitcher plants.

We genotyped wild populations of flowering males and females of three *Nepenthes* taxa using ddRAD-seq, and sequenced a male inflorescence transcriptome. We developed a novel procedure (privacy-rarefaction) to distinguish true sex-specificity from stochastic noise in high-throughput sequencing data.

Results indicated XY-systems in all three *Nepenthes* taxa. The male-specific region of the Y chromosome was much smaller in *Nepenthes* than in *Silene latifolia* (and most other plants) and showed little conservation among taxa, except for the essential pollen development gene DYT1 which was also male-specific in additional taxa. This homomorphic XY sex-determination system likely has a unique origin older than 17.7 My.

Our work characterises the previously unknown sex chromosomes of *Nepenthes* and contributes an innovative, highly sensitive statistical method to detect sex-specific genomic regions in wild populations in general.

## Introduction

### Dioecy and sex chromosomes

Although the majority of flowering plant species have hermaphroditic flowers, plant sexual systems and mechanisms of sex-determination are highly diverse

(Charlesworth 2002; Bachtrog et al. 2014). Only 5-6% of species have female and male flowers on separate individuals (dioecy), while the evolutionary transition to dioecy occurred more than 800 times independently in angiosperms alone (Renner 2014). In contrast to the outcrossing–selfing transition, for which many of the underlying genetic changes have recently been uncovered (Shimizu & Tsuchimatsu 2015), relatively little is known about the transitions from hermaphroditism to dioecy and the mechanisms of sex-determination in plants (Charlesworth 2016). The main hypotheses for the evolution of separate sexes in plants involve conflicting trait optima between the sexual functions, or alternatively, an outcrossing advantage (Charlesworth 1999).

Sex chromosomes are one of the potential determinants of sex. Pairs of sex chromosomes control sex at the individual level, and differ from autosomes mainly in their partial loss of meiotic recombination, and because one of them is limited to one of the sexes (Charlesworth 2016). Some sex chromosome pairs are heteromorphic in karyotypes, while others are homomorphic. Few plant sex-determination systems and sex chromosomes have been studied in detail, and even fewer of these originated independently. This severely limits comparative studies aiming to understand the incidence and stability of sex chromosomes in the tree of life (The Tree of Sex Consortium 2014) and the identification of universal patterns in their evolution and structure. Beyond fundamental evolutionary questions, knowledge of sex-determination systems also has important applications for example in molecular gender phenotyping of juvenile plants in agriculture, plant breeding, and conservation.

This study aimed to develop and apply a novel and robust method to characterise sex-determination systems, unravel the sex-determination system of *Nepenthes* pitcher plants, and investigate biological questions related to the origin of dioecy in this genus.

**Study system**

*Nepenthes* (Nepenthaceae, Caryophyllales) comprises more than 140 taxa of perennial vines and shrubs (Cheek & Jebb 2001, 2013; McPherson 2009) occurring mostly in Southeast Asia. All species are carnivorous plants which supplement their nutrient budget by killing and digesting insects (among other prey), enhancing growth and flowering (Moran & Moran 1998; Pavlovič & Saganová 2015). The complex physiology of carnivory takes place in modified, jug-shaped leaves called pitchers (Juniper et al. 1989; Moran & Clarke 2010).

All *Nepenthes* are dioecious and hermaphrodites are not documented, while the closest relatives, families Ancistrocladaceae, Dioncophyllaceae, Droseraceae, and Drosophyllaceae (Cuénoud et al. 2002; Renner & Specht 2011) are entirely hermaphroditic. Sexual dimorphism in *Nepenthes* could be restricted to the reproductive structures (Kaul 1982), but other traits lack study. Male and female flowers (Figure 1a) are highly diverged because alternative reproductive organs abort early in development (Subramanyam & Narayana 1971). Inflorescences generally share the same structure in males and females (raceme, panicle), but

may differ in tepal colouration and shape, peduncle and rachis length (reviewed in Supporting Information (SI) Table S1 (Clarke 1997, 2001, McPherson 2009, 2011; Clarke et al. 2011); described for only 46 of 138 taxa), and nectar production (Kato 1993; Frazier 2001). Male inflorescences bear in general more flowers (Frazier 2001). Sexual dimorphism in ecology (Barrett & Hough 2013) may exist in *Nepenthes*: In disturbed habitats, adult *N. gracilis* and *N. rafflesiana* were strongly male-biased under open canopy (83% and 100% male individuals, respectively) but were slightly female-biased under closed canopy (34% and 40% males; Frazier 2001), consistent with the hypothesis that males tolerate drier and hotter conditions.

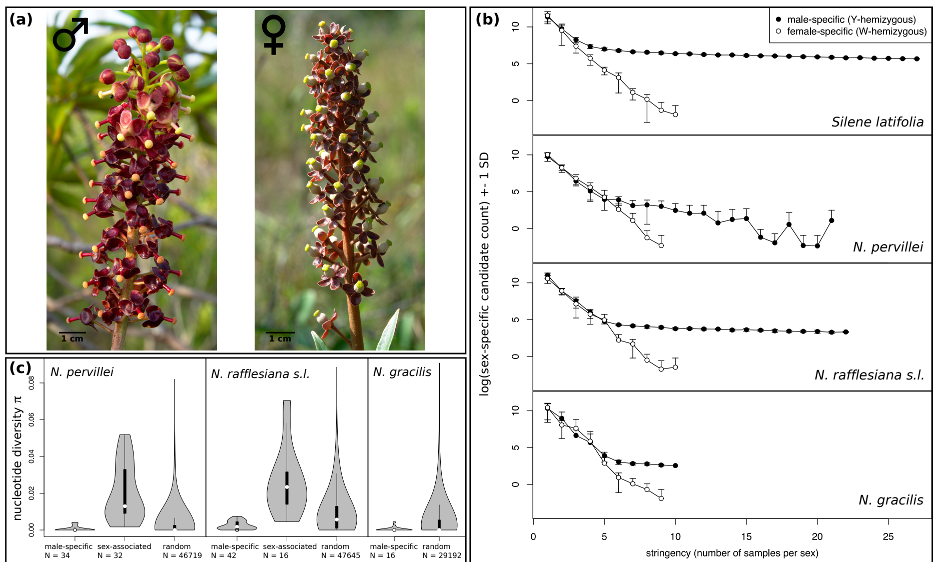



**Figure 1.** (a) male inflorescence of *N. rafflesiana s.l.* (left) and female inflorescence of *N. mirabilis* var. *globosa* (right). (b) Evidence for male-specific markers and an XY sex-determination system in *Silene latifolia* and three *Nepenthes* spp. (privacy-rarefaction curves). Shown are counts of sex-specific markers (y-axis) as a function of the number of samples per sex used to score sex-specificity (x-axis). Sex-specific markers are defined as RAD-tags with mapped reads in all samples from one sex but without any mapped reads in the same number of samples from the other sex. Dots represent averages and whiskers one standard deviation of 200 bootstrapped combinations of males and females. Note natural log-scale of y-axis and hence undefined negative values in SD ranges. (c) Mean per-site nucleotide diversity $\pi$ of RAD-tags in male *Nepenthes* of three taxa for male-specific, sex-associated, and random RAD-tags. All RAD-tags mapping 3-75 reads in $\geq$75% of males per population were included. The same sets of individuals are considered in each category. Sex-associated RAD-tags were absent in *N. gracilis*. Median = white dot, box = 25%–75% quartiles, whiskers = 1.5×interquartile range, violin = estimated kernel density.

The sex-determination mechanism in *Nepenthes* is unknown, but there are no reports of plasticity or reversal of the gender in nature or cultivation (Clarke 2001), suggesting stable sex determination during early development, or a genetic basis. Heteromorphic sex chromosomes are unlikely since a wide range of species share indistinguishably small and uniform chromosomes (2n=80, Heubl & Wistuba 1997).

**Analysis of sex-determination systems**

Cytogenetics and linkage analysis in families are traditional methods to study sex-determination and sex linkage of genes (Charlesworth & Mank 2010). However, these strategies fail in many dioecious organisms because of uninformative karyotypes and prohibitive logistics of breeding. Here we avoid cytology and controlled crosses by instead scanning natural populations for associations between sex and genetic markers.

Two main categories describe molecular genetic sex differences (sex-linkage): sex-association and sex-specificity. Sex-associated loci differ in allele frequency between sexes, and either directly determine sex, for example in polygenic sex-determination systems (reviewed in Bachtrog et al. 2014), or display linkage disequilibrium (LD) with sex-determining loci. Sex-specific loci are private to one sex and entirely absent from the other. They indicate partial divergence between male and female genomes, likely following recombination suppression around sex-determining loci (Charlesworth 1991). Divergence in physical size or structure defines sex chromosomes as heteromorphic as opposed to homomorphic. Male-limitation is referred to as XY-heterogamety and female-limitation as ZW-heterogamety. Sex-specific regions of Y or W chromosomes occur as single copies (haploid, hemizygous) in diploid tissue.

Genomes of both natural and cultivated populations may be scanned for sex differences. Candidate sex-associated or sex-specific loci are expected to display strong mutual LD and reduced genetic diversity relative to autosomal regions. Lack of candidate loci may indicate either non-genetic (environmental) sex-determination, a very small sex-specific region, or multi-locus sex-determination. To date, few studies used population genomics to unravel sex-determination systems and identify sex-linked loci without pedigrees. Chances to find a small sex-linked region with few markers in a large genome are indeed poor. Noteworthy exceptions include the discovery of a sex-specific microsatellite locus in European tree frogs, revealing males as heterogametic in this species (Berset-Brändli et al. 2006). However, marker availability is no longer limiting since the development of restriction site associated DNA sequencing ("RAD-seq" and related methods, Baird et al. 2008; Elshire et al. 2011; Peterson et al. 2012). These methods successfully identified sex-linked markers and sex-determination systems without pedigrees in a Crustacean species (Carmichael et al. 2013) and *Anolis* lizards (Gamble & Zarkower 2014).

Several studies report apparent sex-specific loci in both sexes (Bewick et al. 2013; Gamble & Zarkower 2014; Heikrujam et al. 2015). However, it is generally

expected that there exists only one sex-specific chromosome and only one heterogametic sex. Even in the unusual XYW system (Orzack et al. 1980) or XY-ZW transitional phases (Sander van Doorn & Kirkpatrick 2010), only one sex has a private (dominant) sex chromosome. The paradoxical reports may represent false positives, a problem exacerbated in RAD-seq data with its highly stochastic genotype presence–absence (Mastretta-Yanes et al. 2015). As noted by previous authors (Gamble & Zarkower 2014; Gamble et al. 2015), the rate of false positive sex-specific loci decreases the more males and females are compared, but true positives may be lost at an even higher rate. There exists so far no solution to this quality–quantity trade-off.

#### Aims of the study

We investigated the unknown sex-determination system of dioecious *Nepenthes* pitcher plants and examined whether sex chromosomes across their radiation are derived from a single pair of ancestral autosomes. We developed a statistical procedure that allows distinguishing between true sex-specificity and stochastic absence, and applied this approach to *Nepenthes*. Specifically, we addressed the following questions:

(1) Are there sex-specific and sex-associated RAD-tags in *Nepenthes*?

(2) Do sex-specific and sex-associated RAD-tags conform to theoretical population genetic expectations?

(3) Are sex-specific and sex-associated regions shared among different *Nepenthes* species?

(4) Do these regions contain expressed genes that may contribute to differences between the sexes?

Based on our results we further developed a molecular sexing assay for *Nepenthes*. This tool is suitable for identifying the sex of juveniles and non-flowering adult plants in ecological research, conservation and horticulture.

## Materials and Methods

#### Sampling, ddRAD-seq and genotyping

Samples genotyped by ddRAD-seq (Peterson et al. 2012) and used for this study are shared with those used by Scharmann et al (Chapter I of this thesis), and we refer readers to this publication for specific details. In brief, natural populations of *Nepenthes* were sampled in Brunei Darussalam (Borneo), Singapore, and the island of Mahé, Seychelles. Genomes were Illumina-sequenced in strongly reduced form by focussing on DNA restriction fragments (two enzymes). RAD-tags (contigs) were de novo assembled by clustering reads, followed by mapping, genotype calling and quality filtering. Scans for sex-specific and sex-associated markers were conducted separately on three taxa for which at least 10 morphologically sexed males and females each were available, i.e. *N. pervillei* Blume, *N. gracilis*

Korth., and *N. rafflesiana* sensu lato (defined here as *N. rafflesiana* Jack from Singapore and the Bornean entities *N. rafflesiana* "typical form", *N. rafflesiana* "giant form" (Clarke 1992, 1997; and *N. hemsleyana* Macfarlane; Scharmann & Grafe 2013).

The sex of individual was established on the basis of fresh or dry inflorescences, or else molecularly using a preliminary sexing assay for *N. rafflesiana* s.l. (SI Methods S1). To increase the phylogenetic range of our study and to validate molecular sexing, we furthermore included individuals of known sex for additional species from the field, and species that flowered in cultivation (Table in SI Methods S4). Fresh leaf material was stored in a nucleic acid preserving buffer (Camacho-Sanchez et al. 2013) until further use.

To validate our method for detection of sex-specific markers, we also genotyped several individuals of *Silene latifolia* Poiret, a species with a known XY sex-determination system and heteromorphic sex chromosomes. Details of *Silene* sampling and genotyping are provided in SI Methods S2.

**Detection of sex-specific RAD-tags**

We define sex-specific RAD-tags as mapping sequencing reads from one of the sexes exclusively. The number and identity of sex-specific RAD-tags both carry uncertainties because they depend on the number and identity of male and female individuals compared. We evaluated these uncertainties by resampling methods, separately for each of three *Nepenthes* taxa and *Silene latifolia*. Sex-specificity was tested quantitatively, i.e. for deviation of the observed number of sex-specific RAD-tags from zero, and qualitatively for each RAD-tag.

Unbiased comparisons among data sets with different numbers of individuals and sex ratios were achieved by subsampling male and female individuals such that a 1:1 sex ratio was maintained. To capture the uncertainty over different combinations of individuals, we bootstrapped without replacement for each possible subsample size 200 random sets of males and females. In the quantitative tests, we counted sex-specific RAD-tags for each combination. This generated two distributions, the bootstrapped observed male- and female-specific RAD-tag counts. Then, these two distributions were separately compared to a null distribution derived from permutations of the sexes. The null distribution estimates how many RAD-tags would appear to be sex-specific if the individuals were interchangeable, i.e. if there were no true sex-specific RAD-tags. We calculated a p-value as the proportion of permuted sex-specific counts equal to or larger than the mean of the observed male- resp. female-specific count distribution.

The qualitative test assessed the confidence in sex-specificity for each locus. We again bootstrapped 200 random sets of males and females without replacement from the available individuals, for each possible subsample size. Then, we counted for each locus in how many of the bootstrapped male –female comparisons it emerged as sex-specific. True sex-specific RAD-tags are expected to appear more frequently in such comparisons than false positives whose occurrence is random.

The bootstrap support value for sex-specificity is the count divided by the number of bootstraps. We only reported RAD-tags with 50% or higher bootstrap support.

We named this algorithm privacy-rarefaction and implemented it in a multithreading python script that calls samtools (Li et al. 2009) to read mapping data from .bam alignments (code available on the DVD attached to this thesis, software_code_Scharmann_et_al.tar.gz).

**Detection of sex-associated SNPs**

We conducted $\chi^2$-tests on frequency counts for all bi-allelic SNPs versus sex in PLINK v.1.07 (Purcell et al. 2007). Maximum 25% absent genotypes were tolerated per SNP, and candidate SNPs were accepted as sex-associated at a false discovery rate (Benjamini & Hochberg 1995) smaller or equal to 0.05.

**Population genetics of candidate RAD-tags**

We tested whether LD in male populations differed between sex-specific resp. sex-associated SNPs and the genomic average, represented by 100 randomly selected SNPs. LD ($r^2$) was calculated between but not within RAD-tags using VCFtools v0.1.15 (Danecek et al. 2011). The same contrasts were made for nucleotide diversity $\pi$, which was averaged per RAD-tag using SNPs from .vcf genotypes (VCFtools) while the total number of observed sites per RAD-tag was taken from .bam alignments, applying the same filters to both data (minimum read depth 3, maximum read depth 75, maximum genotype absence 0.25). Significance of differences was evaluated by a randomisation test in R (bootstrap resampling from observed data at equal sample sizes, permutation of observations; p-value = proportion of resampled datasets with difference in means greater or equal to the observed difference; 100k replicates).

**Comparison of candidate RAD-tags to a male inflorescence transcriptome**

We sequenced and assembled the transcriptome of a developing male inflorescence of *Nepenthes khasiana* Hook.f. (SI Methods S3, assembly deposited at Dryad) to identify and annotate sex-linked candidate loci. The transcriptome was searched (a) by BLAST for similarity to candidate RAD-tags (thresholds $\geq$ 90 aligned bases and $\geq$ 75% identity), and (b) by repeating privacy-rarefaction with ddRAD-seq reads directly mapped to the transcriptome rather than the ddRAD-seq assembly (bwa mem (Li 2013), not filtering mapping quality, allowing multiple mappings).

Candidate transcripts from both approaches were annotated by BLAST search against NCBI Genbank nt (version as of Nov 7, 2016), and NCBI Genbank nr (version as of Mar 26, 2016). Transposable elements were detected using RepeatMasker 4.0.6 (Smit et al. 2013) v. 20150807 (Eukaryota). Proteins with at least 50 amino acids were predicted by TransDecoder (Trinity package) and annotated against NCBI Genbank nr, UniProt Swiss-Prot (version as of Aug 17, 2016), and *Arabidopsis thaliana* proteins in UniProtKB (version as of Apr 3, 2016). PFAM domains were detected using hmmer 3.1b1 (Eddy et al.). Database hits were accepted at e-value $\leq 10^{-5}$.

**PCR validation**

Candidate sex-specific RAD-tags were chosen for PCR validation based on a ranking of the highest stringency level reached, bootstrap support, taxonomic overlap, and the quality of annotation of matching transcripts. PCR primers were designed in Geneious R6 (Biomatters Ltd., Auckland, New Zealand). PCR reactions were performed in 15 µl volumes containing 2.5 mM $MgCl_2$, 250 µM of each dNTP, 0.375 units of GoTaq DNA polymerase (Promega, Wisconsin, USA), 1x GoTaq Flexi buffer (Promega), 0.5 µM of each primer, and 1 µl of template DNA (2-20 ng/µl). After initial denaturation for 2 min at 95°C, 30 cycles were run with denaturation at 95°C for 30s, annealing at 50°C for 30s, and extension at 72°C for 1 min, followed by a final extension step of 5 min at 72°C (Labcycler, SensoQuest, Göttingen, Germany). PCR products (2 µl) were separated by electrophoresis in a 2% agarose gel and visualised through fluorescent staining (GelRed, Biotium Inc., Hayward, CA, USA).

## Results

**Sex-specific RAD-tags**

Qualitatively consistent signatures of male-specific RAD-tags were detected independently in *N. pervillei*, *N. gracilis*, *N. rafflesiana* s.l., and *Silene latifolia* (Figure 1b). Numbers of candidate RAD-tags decreased monotonically with subsampling size. This drop in number of shared RAD-tags with increasing number of samples is an inherent and typical property of RAD-seq data, which have large genotype absence caused by a combination of restriction site mutations and stochasticity during library preparation, Illumina sequencing, and bioinformatics (reviewed in Mastretta-Yanes et al. 2015).

The evolutionary plant model for dioecy associated with an XY sex-determination system, *Silene latifolia*, constitutes a positive control for the identification of male-specific RAD-tags since it contains a well-differentiated Y-chromosome. Simultaneously, *S. latifolia* provides a negative control for a ZW-system as evidenced by the decay of false positive female-specific RAD-tags with increasing stringency (subsample size) in all four taxa.

**Sex-associated SNPs**

We detected bi-allelic SNPs associated with the phenotypic sex in *N. pervillei* and *N. rafflesiana* s.l., as well as in *Silene latifolia*, but not in *N. gracilis*. Almost all sex-associated SNPs had an allele frequency close to 0.5 and near-complete heterozygosity in males, but were close to fixation and thus homozygous in females (SI Table S2). The proportion of sex-associated bi-allelic SNPs identified was much lower in *Nepenthes* (*N. pervillei*: 97/38,783=0.25%; *N. rafflesiana* s.l.: 37/222,188=0.017%; *N. gracilis*: 0/50,483=0%) than in *S. latifolia* (2,376/149,311=1.6%).

**LD among sex-specific RAD-tags and sex-associated SNPs**

Sex-specific and sex-associated genomic regions are expected to experience little or no recombination, which should lead to increased LD. Contrary to expectations, LD among sex-specific RAD-tags did not differ from the genomic background

**Table 1.** Estimated relative and physical sizes of the MSY in *S. latifolia* and three *Nepenthes* taxa. Marker numbers are bootstrapped averages at stringency level ten (present in ten males and absent in ten females). Genome sizes were quantified using flow cytometry. Physical size estimates of MSY assumed even marker density along genomes.

| species | Y-specific markers | total markers | genome size 1n [Mbp] | % of markers Y-specific | size of Y-specific region [kbp] |
|---|---|---|---|---|---|
| *Silene latifolia* | 586 | 38455 | 2500 - 3000 | 1.524 | 38000 - 46000 |
| *N. pervillei* | 12 | 51002 | 700 | 0.023 | 161 |
| *N. gracilis* | 13 | 22789 | 850 | 0.057 | 485 |
| *N. rafflesiana s.l.* | 44 | 40509 | 1650 | 0.109 | 1782 |

(p = 0.74) in *N. pervillei*, whereas mean LD among SNPs in sex-associated RAD-tags was elevated by 0.077 units over the genomic background ($p \leq 10^{-5}$). In *N. rafflesiana* s.l., mean pairwise $r^2$ among SNPs located in sex-specific RAD-tags was 0.1 units higher than in the genomic background ($p \leq 10^{-5}$), whereas LD among sex-associated RAD-tags was only slightly higher than the genomic background (c. 0.004 units, p = 0.0011). These tests could not be conducted for *N. gracilis* because no sex-associated RAD-tags were identified and only two sex-specific RAD-tags contained SNPs ($r^2$ = 0.15625).

### Nucleotide diversity of sex-linked RAD-tags

Mean $\pi$ in male-specific RAD-tags tended to be lower than the genomic background in all three taxa (Figure 1c). This difference was significant for *N. rafflesiana* s.l. (p = 0.0005), but not for *N. pervillei* and *N. gracilis* (p = 0.11 and p = 0.065, respectively). In contrast, mean $\pi$ in sex-associated RAD-tags in males was increased over the genomic background for both *N. pervillei* ($p \leq 10^{-5}$, Figure 1c) and *N. rafflesiana* s.l. (p = 0.0043; Figure 1c).

### Size estimates for sex-specific regions

We roughly estimated the relative and absolute sizes of the male-specific regions of Y chromosomes (MSY) using marker counts and genome sizes (Table 1). The data suggest that MSY in *Nepenthes* are 1-2 orders of magnitude smaller than in *S. latifolia*, relative to the species' genome size and in absolute (physical) terms. Within *Nepenthes*, both relative and physical size vary by approximately one order of magnitude.

### Shared candidate loci between species, functional annotations and PCR validation

We recovered six shared candidate sex-specific RAD-tags at stringency level $\geq 5$ in both *N. gracilis* and *N. rafflesiana* s.l., whereas *N. pervillei* shared no candidates with either (SI Table S3). There was no overlap in sex-associated SNPs between the *Nepenthes* species, and no direct overlap between sex-specific RAD-tags and those

with sex-associated SNPs. However, one male-specific RAD-tag of *N. gracilis* and one RAD-tag with sex-associated SNPs of *N. rafflesiana* s.l. both matched with high confidence to the same inflorescence transcript containing a DUF4283 (domain of unknown function, http://pfam.xfam.org/family/PF14111, 09.11.2016).

One male-specific RAD-tag of *N. pervillei* aligned to the transcript of a bHLH transcription factor, and the best matches in all accessed databases were consistently to predicted orthologs of the *Arabidopsis* gene DYSFUNCTIONAL TAPETUM 1 (DYT1). A further sex-associated RAD-tag of *N. pervillei* matched a transcript annotating as *A. thaliana* SEPALLATA-1 (SEP1). This RAD-tag aligned to the predicted 3'-UTR of the putative SEP1-ortholog, and contained two SNPs which were both homozygous in 95% of females and heterozygous in 96% of males. However, further comparisons of putative X–Y divergence of SEP1 were not possible because the male inflorescence transcriptome reads were not heterozygous. In *N. gracilis*, a male-specific RAD-tag matched a long transcript similar to a mitochondrial NADH-ubiquinone oxidoreductase from *Beta vulgaris* (Swiss-Prot). All further candidate loci contained either traces of transposable elements, or no known sequence motifs (SI Table S3).

Complementary to the sex-specificity scan on the ddRAD-de novo reference, we repeated privacy-rarefaction by directly mapping the ddRAD reads to the male inflorescence transcriptome, with the aim to recover further annotated candidate genes. This approach identified seven transcripts as male-specific in at least one species (SI Table S3). We considered only high-confidence male-specific candidate transcripts, reaching at least stringency level four and bootstrap support greater 0.5 in at least one species. No female-specific transcripts (false positives) reached this stringency. A single transcript was male-specific in *N. rafflesiana* s.l. but could not be annotated. Four close transcript "isoforms" (Trinity assembler) were male-specific in both *N. gracilis* and *N. rafflesiana* s.l., but they lacked similarity to any known motif except for one isoform similar to a Jockey-1_Drh retrotransposon. However, two transcripts were male-specific in both *N. pervillei* and *N. rafflesiana* s.l., and one of these also matched a *N. pervillei* male-specific RAD-tag (see above). These two transcripts appear to be close isoforms (putative intron presence–absence), and both annotated as DYT1 (see above).

We tested by PCR whether the putative DYT1-ortholog is male-specific in a broad range of *Nepenthes* species. A single PCR product of approximately 290 bp length was observed exclusively and consistently in phenotypically sexed male *Nepenthes* but never in females (SI Methods S4). We tested multiple males and females (1 vs. 3 to 3 vs. 3) for eight taxa, and 1-2 individuals from 14 further taxa. Presence–absence of the PCR product was fully consistent with the phenotypic sex of all tested individuals (n=56). Sanger sequencing of the PCR product confirmed the identity of the target region. Hence, this locus is male-specific across a phylogenetically broad range of *Nepenthes* species and can be used for molecular sexing of individuals.

## Discussion

### The *Nepenthes* sex-determination system

In natural populations of *Nepenthes*, we discovered both sex-associated markers that were predominantly heterozygous and displayed high nucleotide diversity in males but were mostly homozygous in females, as well as multiple male-specific markers displaying elevated LD and reduced nucleotide diversity. The latter is consistent with theoretical expectations for sex-specific loci, which are hemiploid and have only 1/4 (assuming equal sex ratio) of the effective population size compared to the autosomal genome. Both patterns can arise and persist in interbreeding populations as a consequence of physical linkage to sex-determining loci, (partial) cessation of recombination in the MSY, and X–Y divergence. Together, these findings reveal a genetic basis of sex-determination in *Nepenthes* spp. in which males are the heterogametic sex, i.e. an XY-system. Our interpretations are supported by congruent patterns in the well-known XY-heterogametic *Silene latifolia*.

### Method to extract sex-specific loci from population genotype data

Methods to rapidly genotype individuals across the genome, such as RAPDs, AFLPs, and more recently RNA-seq and ddRAD-seq were repeatedly used to identify sex-specific markers. All these approaches suffer from the common problem that marker absence in some individuals (e.g. due to polymorphism, low coverage or technical artefacts) must be distinguished from true marker absence in the entire sex. Erroneous inference of marker absence in one sex leads to false positive sex-specific markers for the other sex. This likely played a role in studies reporting both male- and female-specific markers (Bewick et al. 2013; Gamble & Zarkower 2014; Heikrujam et al. 2015) in single populations.

Such results are not compatible with theoretical expectations for populations of fully dioecious diploids in panmixia. Under these assumptions, at most one sex, or else none, carries sex-specific alleles or loci as derived from the principles of diploid inheritance (Mendel 1866): If sex is determined by a single locus, it must follow dominant-recessive inheritance (absence constitutes a recessive allele), because co-dominance would produce hermaphrodites or steriles, violating the assumption of a fully dioecious population. Consequently, all loci that are not perfectly physically linked to the sex-determining locus are expected to be shared by both sexes, and sex-specific loci and alleles must be located in a single, non-recombining genomic block that includes a dominant sex-determining allele. If sex is controlled by more than one locus, as in quantitative or polygenic sex-determination, however, by definition no single locus or allele controls sex, and hence all loci and alleles are expected to be shared by both sexes.

We argue that erroneous inference of marker absence, and thus false positive identification of sex-specific markers results from insufficient consideration of uncertainty in presence–absence within and between sexes, which includes sex bias in both sample size and genetic structure of the screened population. We

eliminated this problem through replicated downsampling from a larger pool of observations (individuals) in the same way as rarefaction in community ecology eliminates sampling bias when comparing species richness (Gotelli & Colwell 2001). However, instead of the resampled counts in two groups (habitats), we record the identity and level of sharing (or privacy) between groups (sexes), which is not of interest in conventional rarefaction analysis. As a result, privacy-rarefaction curves, as we name them here, decay rather than increase towards a plateau with increasing sub-sample size (=stringency).

An empirical statistical solution to differentiate between random and true privacy of genomic loci, like the one detailed here, has to our knowledge not been applied before (but see (Kalinowski 2004; Schlüter & Harris 2006; Szpiech et al. 2008) for applications in genetic fingerprinting and diversity estimation). Our method is similar to sex-specific k-mer mapping approaches (Carvalho & Clark 2013; Akagi et al. 2014) but does not require previous knowledge of heterogamety, exhaustive full-genome sequencing, or reference genome assembly. Note that our concept of distinguishing actual absence from stochastic noise is not limited to RAD-seq data. Privacy-rarefaction can be used with any genetic marker, e.g. contigs from transcriptomes and whole-genome resequencing. Furthermore, it complements classical linkage analysis which cannot identify loci in sex-specific regions (Y- resp. W-hemizygous loci) due to lack of recombination.

The power of our method to discover sex-specific markers and heterogamety is limited by the true number of sex-specific loci, the design of population sampling, and the relationship of stochastic noise in genotype presence (consistency of library preparation and read coverage) to the number of individuals per sex. The method performs better the more truly sex-specific loci exist, the higher the read coverage, the more individuals of each sex are included, and the less the population deviates from panmixia (both family structure and geographic structure should be avoided). For example, in a species with XY sex-determination, full-sibling males and females do not contain identical X-chromosomes if their parents carried polymorphic X chromosomes; this may lead to the apparent paradox of finding both male- and female-specific markers. However, this was not the case in our analysis of three *Nepenthes* species and *S. latifolia*, in which female-specific loci were significantly outnumbered by male-specific loci at stringency levels greater than six, and fell to zero above stringency levels ten or eleven (Figure 1b). Moreover, our method is robust to modest levels of erroneous gender phenotyping. We expect that privacy-rarefaction may be useful for diverse analyses.

**Properties of the *Nepenthes* XY system**

*Nepenthes* karyotypes suggest that the sex chromosomes are homomorphic (Heubl & Wistuba 1997), matching the relatively lower proportion of Y-specific and sex-associated RAD-tags of *Nepenthes* compared to *S. latifolia* with its disproportionally large Y-chromosome. Based on theoretical expectations (above), we interpret the occurrence of true sex-specific markers as evidence for a single sex-determining

genomic region. The MSY of *N. pervillei* has an estimated length of only a few hundred kb and thus ranges among the smallest known sex-determining regions in plants, similar to *Vitis* and *Populus* (Filatov 2015).

An expressed gene homologous to *Arabidopsis* DYT1 was found to be male-specific in both *N. pervillei* and *N. rafflesiana* s.l. PCR tests confirmed that this gene is exclusively amplified in males, but never in females, of 22 *Nepenthes* species. We interpret this result as evidence for a conserved core of the *Nepenthes* MSY, given that *N. pervillei* and *N. rafflesiana* s.l. span the largest phylogenetic distance in the genus (Mullins 2000; Meimberg et al. 2001). This shared core MSY further suggests a single common origin of dioecy in *Nepenthes*. Dioecy most likely evolved after the split of dioecious Nepenthaceae and hermaphroditic Droseraceae from their most recent common ancestor (MRCA) at 71.1 (CI 44.2-98.0) Mya, and before the MRCA of extant *Nepenthes* at 17.7 (CI 11.0-24.3) Mya (SI Methods S5).

During the radiation of *Nepenthes*, the MSY has diverged between species. Only six out of 135 male-specific RAD-tags were shared between *N. rafflesiana* s.l. and *N. gracilis*, and none were shared with the more distant *N. pervillei*. Male-specific loci shared between *N. pervillei* and *N. rafflesiana* s.l. were only recovered by directly mapping ddRAD reads to the male inflorescence transcriptome. This suggests that absence of shared sex-specific RAD-tags should not be interpreted as evidence for independent origins of sex chromosomes. Further evidence for a common origin but subsequent interspecific divergence of sex chromosomes is found in a DUF4283-transcript, which was male-specific in *N. gracilis* but sex-associated in *N. rafflesiana* s.l. Apparently, X and MSY alleles (i.e. gametologs) have lost homology (threshold 90% identity) in the former but not in the latter species.

Apart from sequence divergence, variation in absolute and relative abundance of male-specific and sex-associated markers is consistent with variation in the size of the MSY among species, as may be expected in independent evolutionary lineages. Alternatively, limited overlap in male-specific RAD-tags among species outside a core sex-determining region may be compatible with sex chromosome turnover, where the sex-determining region switched into different chromosomal backgrounds (Blaser et al. 2014). Until the genomes of multiple species have been sequenced and compared, this alternative cannot be fully excluded. Yet this aspect does not affect our conclusion of a single common origin of the MSY and dioecy in *Nepenthes*.

**Non-coding DNA and special significance of DYT1 and SEP1**

Of the 41candidate RAD-tags that matched inflorescence transcripts, 36 (88%) could not be annotated or contained transposable elements (TEs). Thus, the large majority of male-specific and sex-associated genomic regions retrieved from *Nepenthes* correspond to non-coding sequences and TEs. Their accumulation is expected in non-recombining regions, and has been reported in species with both heteromorphic (e.g. *Silene*, (Čermák et al. 2008) and homomorphic sex chromosomes (e.g. papaya, Wang et al. 2012). The match of a sex-associated

RAD-tag in *N. gracilis* to a mitochondrial NADH-ubiquinone oxidoreductase was unexpected but may represent either an unspecific match of the RAD-tag (96 bp) to the inflorescence transcript, or else a cyto-nuclear transfer to the sex chromosomes. The occurrence of organellar genes on plant sex chromosomes has been documented in other species (Steflova et al. 2013). erm

Two of the expressed sex-linked genes are of special interest. First, we identified a *Nepenthes* homolog of DYT1 as a genus-wide male-specific (MSY) locus. DYT1 is essential for tapetum development and thus pollen fertility in *A. thaliana* (Zhang et al. 2006), rice (Jung et al. 2005; Wilson & Zhang 2009; Cai et al. 2015), and tomato (Jeong et al. 2014), and we speculate that *Nepenthes* DYT1 is also functionally conserved. Direct validation in *Nepenthes* is currently not feasible due to the lack of transformation protocols and the long generation times. Our analysis suggests that *Nepenthes* DYT1 is absent from females and thus absent from the X chromosome. Such a deletion of DYT1 from the X chromosome would constitute a recessive male-sterility mutation, which is required early in the evolution of dioecy for the transition from a hermaphroditic to a gynodioecious mating system (Charlesworth & Charlesworth 1978).

The second gene of interest is a *Nepenthes* homolog of the homeotic MADS box gene SEP1, an early-acting regulator of floral organ identity in *Arabidopsis* (Pelaz et al. 2000), which was sex-associated in *N. pervillei*. There was near-perfect heterozygosity in two SEP1-linked SNPs in males, whereas these positions were highly homozygous in females, consistent with a location in little- or non-recombining regions of the X and Y chromosomes and gametolog sequence divergence. If SEP1 is functionally conserved in *Nepenthes*, a major component of floral organ development pathways (Theissen et al. 2016) is located on *Nepenthes* sex chromosomes. We thus hypothesize that SEP1 is involved in unisexual flower development in *Nepenthes*. In *Silene latifolia*, however, SEP1 homologs are not directly involved in sex-determination and not located on the sex chromosomes (Matsunaga et al. 2004).

The possible roles of DYT1 and SEP1 in the origin of dioecy in *Nepenthes* require further attention. Even if these genes do not directly determine sex in extant *Nepenthes*, they may have been involved in sexually antagonistic selection during the evolution of dioecy because a deletion of DYT1 or alternative SEP1 alleles might abort non-functional organs at an earlier developmental stage, thus saving resources. However, the completely unisexual floral morphology of extant *Nepenthes* implies that further genetic differences exist between males and females.

**Ecological causes of dioecy in *Nepenthes***

In general, selection for an outcrossing mating system and sexual conflict are hypothesized to drive transitions from hermaphroditism to dioecy (Charlesworth 1999). The *Drosera* species of Western Australia can be compared to *Nepenthes*, as their ancestry, life history and presumably dispersal ability is similar. Most perennial and clonally reproducing Western Australian *Drosera* are self-incompatible and have higher seed abortion than self-compatible annual species (Stace et al.

1997). The mechanisms of self-incompatibility are diverse, suggesting multiple origins or continuing evolution and hence selection for outcrossing due to high genetic load in these poorly dispersing, clonal plants. Ancient hermaphroditic Nepenthaceae may have experienced similar pressure to reduce inbreeding.

Given the extreme nutrient and light limitation common in carnivorous plants (Givnish et al. 1984), sexual conflict may have been rather strong in the hermaphrodite ancestor, assuming that the costs of male and female reproduction differed (Obeso 2002; Barrett & Hough 2013; Zhang et al. 2014). Sexual selection in the form of male–male competition for ovules may favour higher flower numbers in males, whereas costs associated with the development of seeds under nutrient limitation might constrain females to a lower optimal flower number. While male and female reproductive costs and mate competition remain unexplored in carnivorous plants, in *Nepenthes* they may correspond to the difference in flower number between the sexes, which also varies between species (SI Table S1).

**Conclusion**

The discovery of XY sex-determination in Nepenthaceae contributes to a better understanding of the diversity of plant sex determination systems and the molecular and ecological factors associated with dioecy and the evolution of sex chromosomes. As the foundation has been laid, future studies can address sexual conflict, X–Y chromosome divergence, and the identity of sex-determining genes. The species-rich radiation of *Nepenthes* lends itself to comparative studies. With the development of a simple molecular sexing assay, we also provide a tool for future studies on the ecological and physiological correlates of dioecy in *Nepenthes*, and anticipate that future work on *Nepenthes* will benefit from this resource. Our work exemplifies how sex-determination systems of non-model species can be studied and how statistically supported molecular markers suitable for the identification of sex can be developed without the need for prior genetic resources or existing breeding efforts.

## Acknowledgements

We thank N. Zemp for data of *Silene latifolia*. We are indebted to H. Luqman, C. Küffer, J. Mougal, K. Beaver, C. Morel, E. Nancy, A. Street, and T. Kropf for accessing *N. pervillei*. A.H.H. Tinggal and I. Daud with family are thanked for hospitality in Brunei, and L.W. Ngai for assistance with field work in Singapore. The Seychelles Bureau of Standards, the Seychelles Ministry of Environment and Energy, the Ministry of Industry and Primary Resources of Brunei Darussalam, the Agri-Food and Veterinary Authority and the National Parks Board of Singapore kindly granted permits for research, sampling and export of plant material. C. Michel is thanked for preparing sequencing libraries. S. Hartmeyer and U. Zimmermann kindly donated samples of sexed *Nepenthes* from their private collections. Sequencing and computation were carried out with the Genetic Diversity Center Zürich, the Quantitative Genomics Facility Basel, and the Functional Genomics Center Zürich. T. Städler gave helpful comments on the manuscript. This

work was funded by ETH Zürich and partly supported through SNF grant no 31003A_160123. Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies.

## References

Akagi T, Henry IM, Tao R, Comai L (2014) A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science*, **346**, 646-650.

Bachtrog D, Mank JE, Peichel CL et al. (2014) Sex determination: Why so many ways of doing it? *PLOS Biology*, **12**, e1001899.

Baird NA, Etter PD, Atwood TS et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE*, **3**, e3376.

Barrett SCH, Hough J (2013) Sexual dimorphism in flowering plants. *Journal of Experimental Botany*, **64**, 67-82.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289-300.

Berset-Brändli L, Jaquiéry J, Dubey S, Perrin N (2006) A sex-specific marker reveals male heterogamety in European Tree Frogs. *Molecular Biology and Evolution*, **23**, 1104-1106.

Bewick AJ, Chain FJJ, Zimmerman LB et al. (2013) A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biology and Evolution*, **5**, 1087-1098.

Blaser O, Neuenschwander S, Perrin N, Gardner AEA, Day ET (2014) Sex-chromosome turnovers: The hot-potato model. *The American Naturalist*, **183**, 140-146.

Cai C-F, Zhu J, Lou Y et al. (2015) The functional analysis of OsTDF1 reveals a conserved genetic pathway for tapetal development between rice and *Arabidopsis*. *Science Bulletin*, **60**, 1073-1082.

Camacho-Sanchez M, Burraco P, Gomez-Mestre I, Leonard JA (2013) Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, **13**, 663-673.

Carmichael SN, Bekaert M, Taggart JB et al. (2013) Identification of a sex-linked SNP marker in the Salmon Louse (*Lepeophtheirus salmonis*) using RAD sequencing. *PLOS ONE*, **8**, e77832.

Carvalho AB, Clark AG (2013) Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Research*, **23**, 1894-1907.

Čermák T, Kubát Z, Hobza R et al. (2008) Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. *Chromosome Research*, **16**, 961-976.

Charlesworth B (1991) The evolution of sex chromosomes. *Science*, **251**, 1030-1033.

Charlesworth D (1999) Theories of the evolution of dioecy. In: *Gender and Sexual Dimorphism in Flowering Plants* (eds Geber MA, Dawson TE, Delph LF), pp. 33-60. Springer, Berlin, Heidelberg.

Charlesworth D (2002) Plant sex determination and sex chromosomes. *Heredity*, **88**, 94-101.

Charlesworth D (2016) Plant sex chromosomes. *Annual Review of Plant Biology*, **67**, 397-420.

Charlesworth B, Charlesworth D (1978) A model for the evolution of dioecy and gynodioecy. *The American Naturalist*, **112**, 975-997.

Charlesworth D, Mank JE (2010) The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. *Genetics*, **186**, 9-31.

Cheek MR, Jebb MHP (2001) Nepenthaceae. *Flora Malesiana*, **15**, 1-157.

Cheek M, Jebb M (2013) The *Nepenthes micramphora* (Nepenthaceae) group, with two new species from Mindanao, Philippines. *Phytotaxa*, **151**, 25-34.

Clarke CM (1992) *The ecology of metazoan communities in Nepenthes pitcher plants in Borneo, with special reference to the community of Nepenthes bicalcarata Hook. f.* PhD Thesis. University of New England, Armidale, Australia.

Clarke CM (1997) *Nepenthes of Borneo*. Natural History Publications (Borneo), Kota Kinabalu.

Clarke CM (2001) *Nepenthes of Sumatra and Peninsular Malaysia*. Natural History Publications (Borneo), Kota Kinabalu.

Clarke C, Moran JA, Lee CC (2011) *Nepenthes baramensis* (Nepenthaceae) - a new species from north-western Borneo. *Blumea*, **56**, 229-233.

Cuénoud P, Savolainen V, Chatrou LW et al. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *American Journal of Botany*, **89**, 132-144.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Eddy SR, Wheeler TJ, the HMMER development team *HMMER*. http://hmmer.org/.

Elshire RJ, Glaubitz JC, Sun Q et al. (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLOS ONE*, **6**, e19379.

Filatov DA (2015) Homomorphic plant sex chromosomes are coming of age. *Molecular Ecology*, **24**, 3217-3219.

Frazier CK (2001) Pollination and reproductive strategies of *Nepenthes*. In: *Nepenthes of Sumatra and Peninsular Malaysia*, pp. 68-73. Natural History Publications (Borneo), Kota Kinabalu.

Gamble T, Coryell J, Ezaz T et al. (2015) Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Molecular Biology and Evolution*, **32**, 1296-1309.

Gamble T, Zarkower D (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*, **14**, 902-913.

Givnish TJ, Burkhardt EL, Happel RE, Weintraub JD (1984) Carnivory in the bromeliad *Brocchinia reducta*, with a cost/benefit model for the general

restriction of carnivorous plants to sunny, moist, nutrient-poor habitats. *The American Naturalist*, **124**, 479-497.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379-391.

Heikrujam M, Sharma K, Prasad M, Agrawal V (2015) Review on different mechanisms of sex determination and sex-linked molecular markers in dioecious crops: a current update. *Euphytica*, **201**, 161-194.

Heubl G, Wistuba A (1997) A cytological study of the genus *Nepenthes* L. (Nepenthaceae). *Sendtnera*, **4**, 169-172.

Jeong H-J, Kang J-H, Zhao M et al. (2014) Tomato male sterile 1035 is essential for pollen development and meiosis in anthers. *Journal of Experimental Botany*, **65**, 6693-6709.

Jung K-H, Han M-J, Lee Y-S et al. (2005) Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *The Plant Cell*, **17**, 2705-2722.

Juniper BE, Robins RJ, Joel DM (1989) *The carnivorous plants*. Academic Press, London.

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics*, **5**, 539-543.

Kato M (1993) Floral biology of *Nepenthes gracilis* (Nepenthaceae) in Sumatra. *American Journal of Botany*, **80**, 924-927.

Kaul RB (1982) Floral and fruit morphology of *Nepenthes lowii* and *N. villosa*, montane carnivores of Borneo. *American Journal of Botany*, **69**, 793-803.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997 [q-bio].

Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Mastretta-Yanes A, Arrigo N, Alvarez N et al. (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28-41.

Matsunaga S, Uchida W, Kejnovsky E et al. (2004) Characterization of two SEPALLATA MADS-box genes from the dioecious plant *Silene latifolia*. *Sexual Plant Reproduction*, **17**, 189-193.

McPherson S (2009) *Pitcher plants of the Old World*. Redfern Natural History, Poole.

McPherson S (2011) *New Nepenthes - Volume 1*. Redfern Natural History, Poole.

Meimberg H, Wistuba A, Dittrich P, Heubl G (2001) Molecular phylogeny of Nepenthaceae based on cladistic analysis of plastid trnK intron sequence data. *Plant Biology*, **3**, 164-175.

Mendel JG (1866) Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen*, 3-47.

Moran JA, Clarke CM (2010) The carnivorous syndrome in *Nepenthes* pitcher plants. *Plant Signaling & Behavior*, **5**, 644-648.

Moran JA, Moran AJ (1998) Foliar reflectance and vector analysis reveal nutrient stress in prey-deprived pitcher plants (*Nepenthes rafflesiana*). *International Journal of Plant Sciences*, **159**, 996-1001.

Mullins JT (2000) *Molecular Systematics of Nepenthaceae*. PhD Thesis. University of Reading, Reading, UK.

Obeso JR (2002) The costs of reproduction in plants. *New Phytologist*, **155**, 321-348.

Orzack SH, Sohn JJ, Kallman KD, Levin SA, Johnston R (1980) Maintenance of the three sex chromosome polymorphism in the Platyfish, *Xiphophorus maculatus*. *Evolution*, **34**, 663-672.

Pavlovič A, Saganová M (2015) A novel insight into the cost-benefit model for the evolution of botanical carnivory. *Annals of Botany*, **115**, 1075.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, **405**, 200-203.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE*, **7**, e37135.

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559-575.

Renner SS (2014) The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany*.

Renner T, Specht CD (2011) A sticky situation: Assessing adaptations for plant carnivory in the Caryophyllales by means of stochastic character mapping. *International Journal of Plant Sciences*, **172**, 889-901.

Sander van Doorn G, Kirkpatrick M (2010) Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics*, **186**, 629-645.

Scharmann M, Grafe TU (2013) Reinstatement of *Nepenthes hemsleyana* (Nepenthaceae), an endemic pitcher plant from Borneo, with a discussion of associated Nepenthes taxa. *Blumea*, **58**, 8-12.

Schlüter PM, Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. *Molecular Ecology Notes*, **6**, 569-572.

Shimizu KK, Tsuchimatsu T (2015) Evolution of selfing: Recurrent patterns in molecular adaptation. *Annual Review of Ecology, Evolution, and Systematics*, **46**, 593-622.

Smit A, Hubley R, Green P (2013) RepeatMasker Open-4.0. 2013-2016.

Stace HM, Chen L, James SH (1997) Self-incompatibility, seed abortion and clonality in the breeding systems of several Western Australian *Drosera* species (Droseraceae). *Australian Journal of Botany*, **45**, 191-201.

Steflova P, Hobza R, Vyskot B, Kejnovsky E (2013) Strong accumulation of chloroplast DNA in the Y chromosomes of *Rumex acetosa* and *Silene latifolia*. *Cytogenetic and Genome Research*, **142**, 59-65.

Subramanyam K, Narayana LL (1971) A contribution to the floral anatomy of *Nepenthes khasiana* Hook f. *Proceedings of the Indian Academy of Sciences, Section B*, **73**, 124-131.

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498-2504.

The Tree of Sex Consortium (2014) Tree of sex: A database of sexual systems. *Scientific Data*, **1**.

Theißen G, Melzer R, Rümpler F (2016) MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development*, **143**, 3259-3271.

Wang J, Na J-K, Yu Q et al. (2012) Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences*, **109**, 13710-13715.

Wilson ZA, Zhang D-B (2009) From *Arabidopsis* to rice: pathways in pollen development. *Journal of Experimental Botany*, **60**, 1479-1492.

Zhang S, Jiang H, Zhao H, Korpelainen H, Li C (2014) Sexually different physiological responses of *Populus cathayana* to nitrogen and phosphorus deficiencies. *Tree Physiology*, **34**, 343-354.

Zhang W, Sun Y, Timofejeva L et al. (2006) Regulation of *Arabidopsis* tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development*, **133**, 3085-3095.

## Author contributions

MS performed the research including data collection and analysis; TUG and FM contributed logistic support and materials; MS and AW designed and interpreted the research and wrote the manuscript.

# Chapter 2 Supporting Information Methods S1, S2, S3, S4, S5

---

further Supporting Information is available on the CD attached to this thesis

## Methods S1 Preliminary molecular sexing assay for *Nepenthes rafflesiana s.l.*

An initial sequencing library contained a sufficient number of sexed individuals for *Nepenthes rafflesiana s.l.* and *N. gracilis* from Borneo. Collection and DNA extraction are detailed in the main text. We commissioned the Genomic Diversity Facility (Cornell University, Ithaca, NY, USA) with library construction and sequencing, following the GBS protocol (Elshire et al. 2011). After optimisation, the restriction enzyme Sfb1 was chosen and the library was sequenced for 100-bp single-end reads in two Illumina HiSeq lanes.

At this early stage, we employed a simpler version of the resampling approach to detect sex-specific loci, using the Stacks pipeline (Catchen et al. 2013) instead of the dDocent approach (Puritz et al. 2014) of genotyping. The populations module of Stacks was iterated over different combinations of real and permuted males and females, thereby revealing numbers and identities of likely sex-specific loci following the same logic as described in the main text ("privacy-rarefaction"). This attempt was successful in *N. rafflesiana* "typical form" (Borneo), but failed to identify any sex-specific loci in *N. gracilis*. We took the top 10 best candidate loci for *N. rafflesiana* "typical form" (Borneo) and designed PCR primers for validation of sex-specificity (PCR conditions as described in the main text). Two of these loci amplified from males exclusively (private gel band at expected size), as verified in all males that were used for the genotyping and several further samples that had not been used previously. The same markers also amplified specifically from known males but not females of *N. hemsleyana* and *N. rafflesiana* "giant form" (Borneo). However, these markers were unspecific for all other tested species (*N. ampullaria*, *N. bicalcarata*, *N. gracilis*, *N. mirabilis*). We consequently used these two markers to molecularly sex additional individuals of *N. rafflesiana* "typical form" (Borneo) and *N. hemsleyana* that were included in the later, full sample set genotyped with ddRAD-seq. To conclude, the sex of most *N. hemsleyana* and several of the *N. rafflesiana* "typical form" (Borneo) were determined not on the phenotype but molecularly with markers developed through this initial GBS dataset.

## Methods S2 Genotyping of *Silene latifolia*

Wild populations of *Silene latifolia* were sampled across Switzerland and leaves preserved by drying in silica gel. The phenotypic sex of individuals was recorded.

After DNA extraction (Qiagen DNeasy Plant Mini Kit), a set of 95 samples was commissioned for library construction to the Genomic Diversity Facility (Cornell University, Ithaca, NY, USA), following the GBS protocol (Elshire et al. 2011). The restriction enzyme ApeKI was chosen after optimisation. This library was sequenced in two lanes of an Illumina HiSeq to ensure sufficient coverage.

The bioinformatics for *Silene* genotyping were identical to those employed for *Nepenthes*, as outlined in the main text respectively in Scharmann et al (chapter I of this thesis), i.e. de novo reference assembly following a modified dDocent pipeline (Puritz et al. 2014), read mapping, variant calling and quality filtering.

## Methods S3 Male inflorescence transcriptome

Plants of *Nepenthes khasiana* (in vitro propagated material from Borneo Exotics (Pvt) Ltd., Sri Lanka) were grown in a greenhouse where they flowered regularly. For the transcriptome of a male inflorescence of *N. khasiana* (length c. 2 cm, many buds of c. 1-3 mm diameter each), we extracted RNA using the Total RNA Mini Kit (Plant) (Geneaid Biotech Ltd, New Taipei City, Taiwan) with the "PRB" lysis buffer, which yielded undegraded high quality RNA (RIN 7.1; Plant RNA Nano Assay, Agilent Bioanalyzer). A cDNA library was generated (NEBNext Ultra Directional RNA Library Prep Kit for Illumina, New England Biolabs, Ipswich MA, USA) and sequenced in one lane of the Illumina MiSeq for 150 bp paired-end reads (GDC ETHZ). A total of 18.7 million PE reads were obtained, and a reference transcriptome was de novo assembled by the Trinity pipeline (Grabherr et al. 2011). Best ORFs were extracted by the Transdecoder script.

## Methods S4 A molecular sexing assay for the genus *Nepenthes*

Based on the evidence for male-specific genomic regions (non-recombing Y-chromosomeal region), we developed an assay to sex *Nepenthes* molecularly. Here we test it with phenotypically sexed individuals from 22 different *Nepenthes* spp. (Methods S4 Table S4-1). The assay is likely applicable to further *Nepenthes* spp., but we recommend to validate it using several phenotypically sexed individuals before application to a novel species.

### DNA extraction

*Nepenthes* tissue contains substances that strongly inhibit PCR, as simple extraction protocols without purification steps (as suggested by Hobza & Widmer (2008) did not yield any amplification. We thus used the silica-column kit NucleoSpin Plant II from Macherey Nagel (Düren, Germany). For optimal yield, the tissue was completely powdered before the lysation step. To achieve this, tissue was flash-frozen in liquid nitrogen and then crushed in disposable, folded paper envelopes using pliers. The resulting coarse powder had to be still frozen, and was transferred to a 2 ml cryotube (Sarstedt No. 72.694.005, Screw Cap Micro Tube, 2 ml, PP, conical and skirted base) with three steel beads. On a shaker mill, cycles of shaking (up to 15 s) and flash-freezing in a liquid nitrogen bath were repeated until the material was a fine dust. Acceleration on the shaker

**Supporting Information Methods S4 Table S4-1.** *Nepenthes* spp. with phenotypically verified sex used for broader taxonomic validation of a male-specific PCR marker

| *Nepenthes* species | N male | N female | source |
|---|---|---|---|
| *adnata* | 1 | 1 | cultivated |
| *albomarginata* | 1 | 3 | wild populations, cultivated |
| *ampullaria* | 2 | 2 | wild populations |
| *bicalcarata* | 2 | 2 | wild populations |
| *clipeata* | 1 | - | cultivated |
| *gracilis* | 3 | 3 | wild populations, cultivated |
| *hemsleyana* | 2 | 2 | wild populations |
| *khasiana* | 1 | - | cultivated |
| *maxima* | 2 | - | cultivated |
| *mira* | - | 1 | cultivated |
| *mirabilis* | 3 | 3 | wild populations |
| *mirabilis* var. *globosa* | - | 1 | cultivated |
| *pervillei* | 3 | 3 | wild populations |
| *petiolata* | - | 1 | cultivated |
| *rafflesiana* typical form Borneo | 3 | 3 | wild populations |
| *singalana* | 1 | - | cultivated |
| *talangensis* | 1 | - | cultivated |
| *tentaculata* | 1 | - | cultivated |
| *truncata* | 1 | - | cultivated |
| *veitchii* | - | 1 | cultivated |
| *ventricosa* | - | 1 | cultivated |
| x *trusmadiensis* | 1 | - | cultivated |

mill was carefully adjusted to the maximum possible level that did not break the frozen cryotubes. Lysis buffer was added directly to the tissue dust without prior thawing. All other steps followed the kit instructions.

**PCR amplification of control and sex-specific sequences**

The assay involves four primers: one pair targets a male-specific region within the putative *Nepenthes* ortholog of the *Arabidopsis thaliana* DYT1 gene:

25100_L96_2_F: 5′-AATTCACTGATTCGGATCACG-3′

25100_L96_294_R: 5′-CGATCGCGTCGCAAAGTATG-3′

while the other targets a sequence that is common to both sexes, the mitochondrial cox1 gene (Cho et al. 1998):

IP53: 5′-GGAGGAGTTGATTTAGC-3′

cox1.6KR: 5′-AAGGCTGGAGGGCTTTGTAC-3′

As suggested by Hobza & Widmer (2008), the common target is used as an internal control of each reaction. It ensures that poor template quality or other technical issues are recognised as such, instead of confusion with true absence of the sex-specific target, i.e. the expectation for females. The two regions can also be amplified in separate reactions.

The reaction is performed in 15 µl volumes containing 2.5 mM $MgCl_2$, 250 µM of each dNTP, 0.375 units of GoTaq DNA polymerase (Promega, Wisconsin, USA), 1x GoTaq Flexi buffer (Promega), 0.5 µM of each of the four primers, and 1 µl of template DNA extract (c. 5-20 ng/µl). After initial denaturation for 2 min at 95°C, 30 cycles are run of denaturation for 30s at 95°C, annealing for 30s at 50°C and extension for 1 min at 72°C, followed by a final extension step of 5 min at 72°C (Thermocycler, e.g. Labcycler, SensoQuest, Göttingen, Germany).

**visualisation and scoring**
PCR products are separated in 2% agarose gels and visualised by fluorescent dye (Methods S4 Fig. S4-1). Successful assays contain at least one strong band at 600-700 bp length, corresponding to the control cox1 fragment. The presence of a strong band at c. 290 bp characterises male individuals, while females do not contain this band. Several other, weaker bands of different length may be present. These are likely unspecific products of the control primer pair, as we could never observe them when applying the sexing primer pair exclusively.

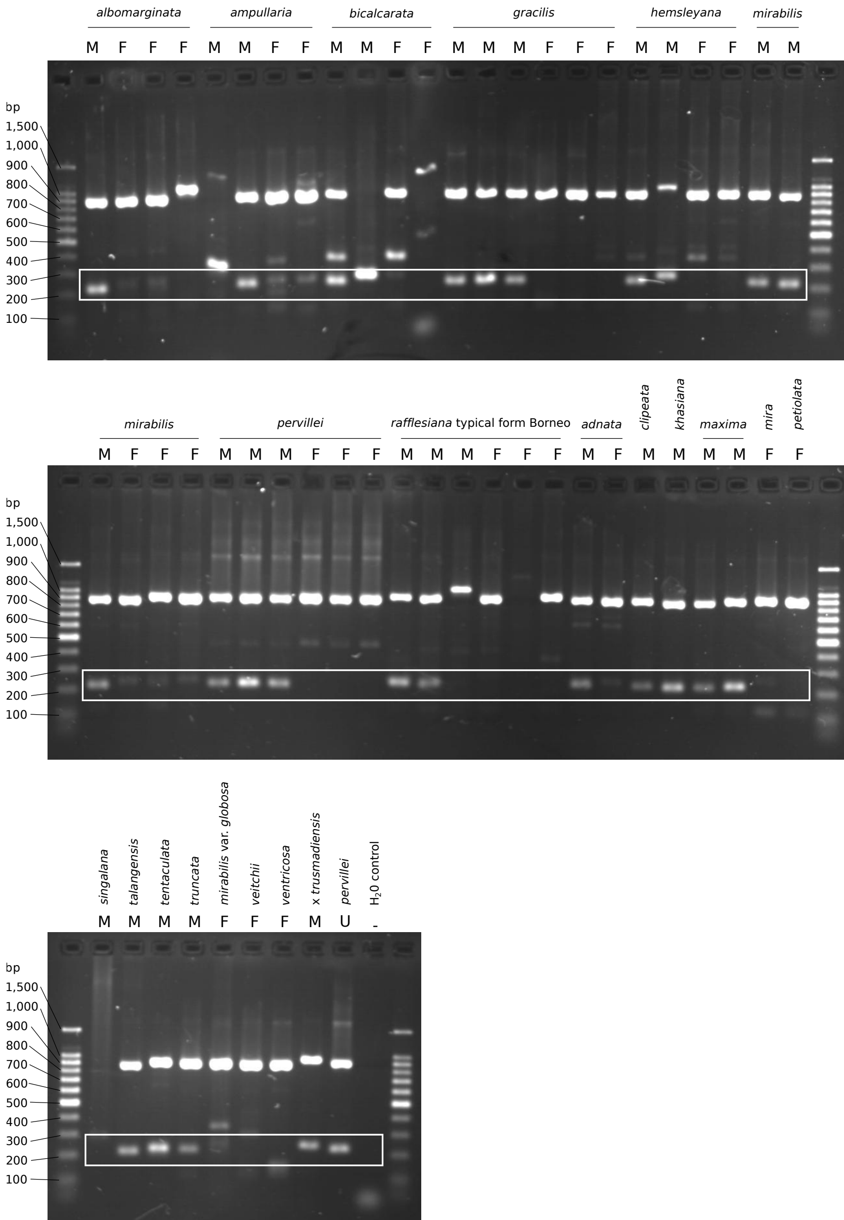
## Methods S5 Phylogenetic dating of *Nepenthes*

### Materials and method

We dated the genus *Nepenthes* by combining new transcriptome data for *Nepenthes* with previously published transcriptome data of the Venus flytrap (*Dionaea muscipula*, Bemm et al. 2016), a transcriptome-based phylogeny of Caryophyllales (Yang et al. 2015), and dates of Angiosperm diversification (Magallón et al. 2015). *Dionaea* represents Droseraceae, which is among the closest living sister lineages of *Nepenthes* (Brockington et al. 2009; Soltis et al. 2011).

In the first step, transcriptomes of 12 *Nepenthes* spp. (Scharmann et al in preparation; Chapter 3 of this thesis) were assembled de-novo using Trinity (Grabherr et al. 2011). The raw assembly for *Dionaea* (v1.03) was downloaded from http://tbro.carnivorom.com/. We extracted candidate ORFs resp. peptide sequences with TransDecoder.LongOrfs v3.0.0 and TransDecoder.Predict. To reduce the sequence collections even further in a meaningful way, we retained only peptides that were similar (e-value $\leq 10^{-5}$) to any gene from all available Eudicot plant genome assemblies (NCBI Genbank, accessed 6 June 2016).

In the second step, we emended a taxon-subset of the peptide sequence matrix for the 1,122 genes of Yang et al. (2015) with orthologs from the *Dionaea* and *Nepenthes* transcriptomes. A custom python script was used to decompose the matrix by gene and taxon using the also available gene model file. The peptide sequences of "*Nepenthes alata* (WQUF)" were used to identify orthologs in the new transcriptomes by reciprocal best hit (blastp) with an e-value cutoff of 0.01. Third, a matrix was re-assembled with 21 of the original taxa and 13 newly added taxa, by globally re-aligning all ortholog peptide sequences of each gene using MUSCLE (Edgar 2004), and concatenation of the alignments. A new gene model file was generated in the process to allow partitioned analysis of the matrix. The new alignment was slightly longer than the original (550,076 instead of 504,850




**Supporting Information Methods S4 Fig. S4-1.** Electrophoretic gel confirming molecular sexing of *Nepenthes* spp. listed in Methods S4 Table S4-1. The region for male-specific bands is highlighted by the white box.

sites), contained 34 taxa, and 21.7% gap characters. The 13 taxa we added showed very high sequence occupancy, each containing >1,000 of the original 1,122 genes of Yang et al. (2015).

The maximum likelihood tree was reconstructed with the same method and partitioned by genes as before (RAxML -m PROTCATWAG -q ; Yang et al. 2015). SH-like support was calculated using RAxML -f J option.

We then dated the divergence times on a pruned version of this tree (see below) using the RelTime algorithm (Tamura et al. 2012) as implemented in MEGA-CC (Kumar et al. 2012). RelTime is a non-Bayesian method for dating of phylogenentic trees that produces estimates similar to those from e.g. BEAST and MCMCtree, but it is orders of magnitudes faster and thus copes with genomics-scale alignments (Mello et al. 2017). The pruned tree contained only Brassicaceae (*Arabidopsis thaliana*) as the outgroup, and hence the alignment given to RelTime was also reduced with the same method as above (32 taxa, 550,360 sites, 22% gaps). We specified the WAG substitution model with 5 gamma-distributed rate categories and invariant sites. For 13 nodes that were also present in the Angiosperm time-tree of Magallón et al. (2015), we supplied absolute time calibrations in the form of upper and lower limits on age (Methods S5 Table S5-1).

Preliminary runs of RAxML and RelTime revealed that inclusion of Fabaceae, Rosaceae and Brassicaceae resulted in the same topology as retrieved by Yang et al. (2015), but enforcing these "Rosids" as a monophyletic outgroup caused RelTime to fit negative branch lengths near the root. However, reducing the outgroup to just Brassicaceae, RAxML found a rather different topology compared to Yang et al. (2015), and to this tree RelTime fitted negative branch lengths among the major lineages of Caryophyllales. Thus, to avoid biologically not interpretable negative branch lengths, we obtained a topology using the three outgroup taxa, but pruned this tree and the alignment to retain only Brassicaceae during RelTime dating.

**Results and discussion**

We retrieved largely the same topology as Yang et al. (2015) for our subset of Caryophyllales taxa, with full SH-like LRT support for all nodes (tree not shown). The only exception, a grouping of Sarcobataceae as sister to Nyctaginaceae instead of Phytoloccaceae, occured in a lineage distant to *Nepenthes*. *Nepenthes* was monophyletic and grouped as sister to Droseraceae (*Dionaea muscipula*). This carnivorous lineage was sister to Frankeniaceae-Plumbaginaceae-Polygonaceae as reported before (Yang et al. 2015). The stem age of *Nepenthes* was estimated at 71.1 (CI 44.2 - 98.0) Mya, when it split from its presumed sister Droseraceae. The crown of *Nepenthes* is marked by the most basal species *N. pervillei*, and estimated here at 17.7 (CI 11.0 - 24.3) Mya.

However, we interpret these time estimates with great caution. First, the identity of the closest living relative of *Nepenthes* has not yet reached a consensus. Candidates are the lineage of sticky-leaf carnivores *Drosophyllum* and *Triphyophyllum* and several tropical lianas that appear to have lost carnivory secondarily (Heubl

et al. 2006; Renner & Specht 2011), and the Droseraceae (Brockington et al. 2009; Soltis et al. 2011), or *Nepenthes* may be part of a deep reticulation event (Walker et al. 2017). We focussed on the Droseraceae because this was the only family for which transcriptome data was available.

Second, the divergence times that we took from the literature (secondary calibrations) may change in the future, as these were based on few genetic loci, fossils may be re-interpreted, and estimation methods change.

All previous attempts of molecular dating in *Nepenthes* (Meimberg 2002; Merckx et al. 2015) involved a presumed *Nepenthes* pollen fossil from the European Eocene (c. 50 million years ago, Krutzsch 1985). However, the attribution of this fossil to an ancestor of recent *Nepenthes* is not justified – it is larger than recent *Nepenthes* pollen but instead fits in the range of Droseraceae (Cheek & Jebb 2001). Thus, Krutzsch's pollen fossils are at best indicative of the European Eocene presence of some lineage with Droseraceae-Nepenthaceae affinity but do not imply an age for modern *Nepenthes*.

**Supporting Information Methods S5 Table S5-1.** Absolute time calibrations as constraining upper and lower boundaries for the RelTime analysis, taken from Magallón et al. (2015). These are 95% confidence limits for the age in million years of the most recent common ancestor (MRCA) of 13 pairs of plant families studied by both Yang et al. (2015) and Magallón et al. (2015).

| MRCA of | min time | max time |
|---|---|---|
| Cactaceae – Portulacaceae | 15.08 | 48.15 |
| Cactaceae – Talinaceae | 18.75 | 53.24 |
| Caryophyllaceae – Amaranthaceae | 50.34 | 88.56 |
| Cactaceae – Molluginaceae | 54.22 | 84.99 |
| Phytolaccaceae – Sarcobataceae | 63.24 | 73.41 |
| Plumbaginaceae – Polygonaceae | 65.63 | 78.21 |
| Aizoaceae – Nyctaginaceae | 72.5 | 77.87 |
| Phrymaceae – Solanaceae | 77 | 103 |
| Frankeniaceae – Plumbaginaceae | 83 | 101 |
| Caryophyllaceae – Physenaceae | 90 | 101 |
| Caryophyllaceae – Simmondsiaceae | 95.44 | 105.53 |
| Caryophyllaceae – Solanaceae | 114 | 123 |
| Brassicaceae – Nyctaginaceae | 120.87 | 126.49 |

## References

Bemm F, Becker D, Larisch C et al. (2016) Venus flytrap carnivorous lifestyle builds on herbivore defense strategies. *Genome Research*, **26**, 812-825.

Brockington SF, Alexandre R, Ramdial J et al. (2009) Phylogeny of the Caryophyllales sensu lato: Revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. *International Journal of Plant Sciences*, **170**, 627-643.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124-3140.

Cheek MR, Jebb MHP (2001) Nepenthaceae. *Flora Malesiana*, **15**, 1-157.

Cho Y, Qiu Y-L, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences*, **95**, 14244-14249.

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Elshire RJ, Glaubitz JC, Sun Q et al. (2011) A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLOS ONE*, **6**, e19379.

Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-652.

Heubl G, Bringmann G, Meimberg H (2006) Molecular phylogeny and character evolution of carnivorous plant families in Caryophyllales - revisited. *Plant Biology*, **8**, 821-830.

Hobza R, Widmer A (2008) Efficient molecular sexing in dioecious *Silene latifolia* and *S. dioica* and paternity analysis in F1 hybrids. *Molecular Ecology Resources*, **8**, 1274-1276.

Krutzsch W (1985) über *Nepenthes*-Pollen (alias ?*Droseridites*? p.p.) im europäischen Tertiär. *Gleditschia*, **13**, 89-93.

Kumar S, Stecher G, Peterson D, Tamura K (2012) MEGA-CC: Computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*, **28**, 2685-2686.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T (2015) A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, **207**, 437-453.

Meimberg H (2002) *Molekular-Systematische Untersuchungen an den Familien Nepenthaceae und Ancistrocladaceae sowie verwandter Taxa aus der Unterklasse Caryophyllidae s. l.* PhD Thesis. Ludwig-Maximilians-Universität München, München, Germany.

Mello B, Tao Q, Tamura K, Kumar S (2017) Fast and accurate estimates of divergence times from big data. *Molecular Biology and Evolution*, **34**, 45-50.

Merckx VSFT, Hendriks KP, Beentjes KK et al. (2015) Evolution of endemism on a young tropical mountain. *Nature*, **524**, 347-350.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e431.

Renner T, Specht CD (2011) A sticky situation: Assessing adaptations for plant carnivory in the Caryophyllales by means of stochastic character mapping. *International Journal of Plant Sciences*, **172**, 889-901.

Soltis DE, Smith SA, Cellinese N et al. (2011) Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, **98**, 704-730.

Tamura K, Battistuzzi FU, Billing-Ross P et al. (2012) Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences*, **109**, 19333-19338.

Walker JF, Yang Y, Moore MJ et al. (2017) Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *bioRxiv*, 115741.

Yang Y, Moore MJ, Brockington SF et al. (2015) Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution*, **32**, 2001-2014.

# Chapter 3
## Adaptive evolution of carnivory in *Nepenthes* pitcher plants: a comparative transcriptomics and proteomics perspective

an unpublished manuscript co-authored by

M Scharmann[1], F Metali[2], T U Grafe[2], A Widmer[1]

[1] ETH Zurich, Plant Ecological Genetics, Institute of Integrative Biology (IBZ), Universitätstrasse 16, 8092 Zurich, Switzerland
[2] Faculty of Science, Universiti Brunei Darussalam, Jalan Tungku Link, Bandar Seri Begawan, BE1410 Brunei Darussalam

### Abstract

While phenotypic diversity associated with evolutionary radiations is easily recognized, the function of specific traits and their adaptive relevance often remain elusive. Here we ask whether adaptive evolution has shaped genes underlying carnivory in the radiation of *Nepenthes*, an enigmatic genus of pitcher plants encompassing approximately 150 ecologically and morphologically highly diverse species. We investigate 13 species representing a wide range of nutrient acquisition strategies and trap physiologies, as well as geographic and climatic origins. Using a combination of proteomics and transcriptomics we identified hundreds of secreted pitcher fluid proteins (PFPs), reflecting previously unknown intra- and interspecific functional diversity. Further genes related to plant carnivory were uncovered by experimentally feeding *Nepenthes* traps and analyses of gene expression changes, which hinted towards a strong physiological shift associated with feeding, away from photosynthesis towards heterotrophy, proteolysis and protein synthesis. We found that many thousands of genes show adaptations in protein sequences and regulatory divergence during the *Nepenthes* radiation, but PFPs and carnivory-related genes were in comparison to other genes exceedingly common among them. This bias seemed to be already present during the early stage of divergence, as carnivory-related genes were over-represented among differentially expressed genes in a young species pair, and their pitcher traps diverged overall more in gene expression than the leaves. Our results suggest that the molecular basis of the carnivorous syndrome was disproportionally targeted by positive selection during the *Nepenthes* radiation. This study demonstrates how

the evolutionary relevance of suspect key traits can be investigated at the sequence and expression level of underlying genes, which are increasingly accessible for any organism.

## Introduction

Carnivorous plants have long attracted the interest of scientists (e.g. Hooker 1874; Darwin 1875). In recent years, a solid understanding of the physiology, ecology and evolution of this enigmatic nutrient acquisition strategy has been developed (Juniper et al. 1989; Ellison & Gotelli 2009; Krol et al. 2012; Fukushima et al. 2017; Lan et al. 2017) but we currently know very little about the subsequent diversification of carnivorous plant lineages. Carnivory has repeatedly and independently evolved across the phylogeny of angiosperms and several, but not all (Krol et al. 2012; Pereira et al. 2012), carnivorous plant genera have undergone evolutionary radiations. The role that carnivory has played in such radiations remains unstudied. Here we raise the question whether the carnivorous syndrome itself underwent adaptive evolution and has potentially contributed to rapid diversification, which is thought to derive from a combination of both adaptive and non-adaptive processes (Losos & Mahler 2010; Givnish 2015). Alternatively, carnivory could be a conserved trait or evolve neutrally once established and have no effect on the subsequent evolution of the plant lineage.

Several frameworks exist to test whether specific traits have played important roles in evolutionary radiations. Such traits, often referred to as key adaptation traits, contribute disproportionally to functional diversity and are repeatedly targeted by diversifying selection during evolutionary radiations (e.g. flower colour and spur length in *Aquilegia*, (Whittall & Hodges 2007; Hodges & Derieg 2009). Among the most popular approaches are tests of macro-evolutionary models that aim to relate the origin of particular traits to shifts in phylogenetic diversification rates (e.g. Sanderson & Donoghue 1996; Rabosky 2014; Käfer & Mousset 2014). A novel and more direct approach applies genome-wide scans for molecular signatures of diversifying selection across radiations (e.g. Roux et al. 2014; Brawand et al. 2014; Pease et al. 2016; Nevado et al. 2016). Signatures of past selection episodes can be inferred from DNA and protein sequences by analysing ratios of non-synonymous to synonymous substitutions in codon alignments (Nielsen 2005; Yang 2014a; Smith et al. 2015). In addition to amino acid substitutions, regulatory changes have long been hypothesised to play important roles in adaptation (King & Wilson 1975). Recent developments now allow better inference of gene expression evolution levels through non-parametric and model-based frameworks of quantitative trait evolution (Bedford & Hartl 2009; Rohlfs et al. 2014; Roux et al. 2015; Rohlfs & Nielsen 2015), and new empirical results have established that regulatory changes can lead to adaptation (Fraser 2011; Enard et al. 2014; Naranjo et al. 2015). The contributions to adaptation of gene expression changes relative to coding sequence (structural) changes, however, remains an important question (Hoekstra & Coyne 2007; Fraser 2011; Romero et

al. 2012). Comparative datasets encompassing multiple species across a radiation and covering representative samples of organismal genomes or transcriptomes hold the potential to reveal the molecular basis of adaptations and to answer questions on the contributions of gene sequence and gene expression changes to adaptation.

Plant carnivory is a complex quantitative trait whose genetic basis is largely unknown. Gene functions cannot directly be inferred from model species because they lack this trait. Nevertheless, carnivory is well suited for explorations of some of the underlying genes using a combination of proteomics and transcriptomics (e.g. Voelckel et al. 2017). Prey digestion in carnivorous plants is a physiological trait for which direct links can be established between the digestive enzymes, the genes encoding these enzymes, and their role in carnivory. Most available information on enzymes used by carnivorous plants to digest their prey comes from *Nepenthes* (Nepenthaceae, Caryophyllales). As early as 1874, it was noted that the fluid of *Nepenthes* pitcher plants degrades meat, egg white and cartilage (Hooker 1874), but digestive enzymes and their secretion were characterised only in the second half of the 20th century (Juniper et al. 1989). A major debate ensued concerning the origin of these enzymes from either plant or microbes living in the pitcher fluid (Frazier 2000). While it has been established that microbes contribute to the enzymatic activity and nitrogen release in natural pitcher fluid (Takeuchi et al. 2011; Lam et al. 2017), digestive enzymes are present and active even in sterile pitchers (Juniper et al. 1989; Buch et al. 2013). Evidence for plant-derived digestive enzymes was corroborated by sequencing the most abundant group of proteins found in pitcher fluids, aspartic peptidases called Nepenthesins (Athauda et al. 1998), and their subsequent cloning from the plant genome (Athauda et al. 2004).

The combination of proteomic and genomic techniques has since verified that *Nepenthes* secrete into their pitcher fluid a host of hydrolytic and antimicrobial proteins (Eilenberg et al. 2006; Hatano & Hamada 2008, 2012; Rottloff et al. 2016; Lee et al. 2016; Fukushima et al. 2017), which we collectively refer to as pitcher fluid proteins (PFPs). As they are directly degrading prey and potentially control pitcher microbial communities, PFPs are key elements of plant carnivory. At least some PFPs are not constitutively produced but are induced after prey capture (Buch et al. 2015; Yilamujiang et al. 2016) through upregulation of gene expression in the glandular, submerged pitcher wall that is in direct physical contact with the pitcher fluid (An et al. 2001, 2002; Eilenberg et al. 2006; Rottloff et al. 2011). These observations suggest that sensory mechanisms detect prey and that PFP synthesis may be energetically expensive for the plant.

The genus *Nepenthes* (Nepenthaceae, Caryophyllales) is highly diverse and encompasses approximately 150 species distributed mainly in Southeast Asia (Cheek & Jebb 2001; McPherson 2009; Moran & Clarke 2010). Although all *Nepenthes* produce pitcher traps, there is a large diversity in morphology, biomechanics, and chemistry of the digestive fluid (Cheek & Jebb 2001; McPherson 2009; Moran &

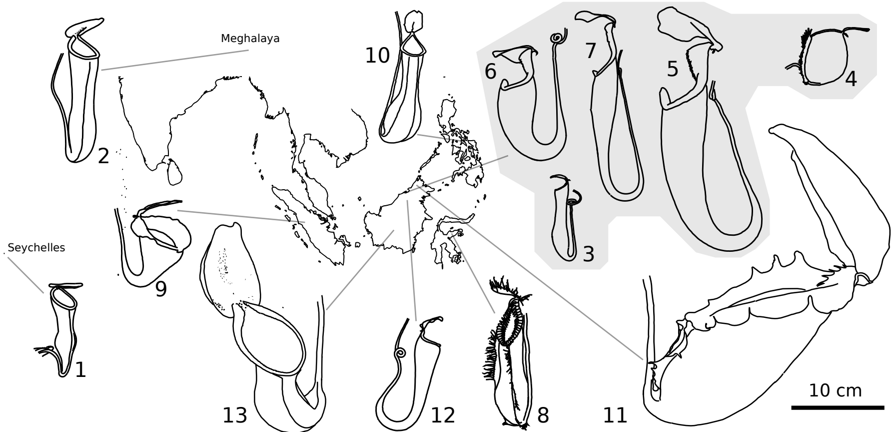
Clarke 2010) that corresponds in part to differences in prey spectra. *Nepenthes* include insect-eating generalists, a termite specialist, a leaf-litter specialist, and several coprophagous species that attract small vertebrates and intercept their excreta (Moran & Clarke 2010; Schöner et al. 2017). To what extent pitcher trap diversity is also reflected at the physiological level of prey digestion remains an open question. To date, only two studies have suggested that variation in proteinase activity and PFP composition exists among *Nepenthes* species (Takahashi et al. 2007; Biteau et al. 2013).

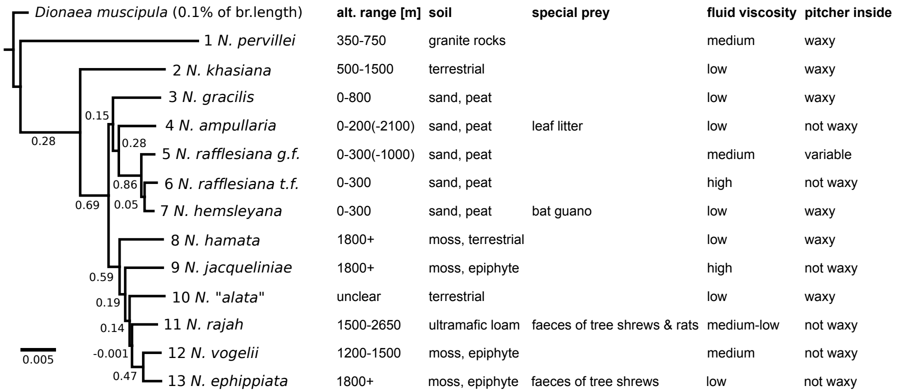
Here we combine proteomics and transcriptomics to investigate the genetic basis of carnivory in *Nepenthes* pitcher plants. We explore the diversity of PFPs and changes in gene expression that accompany feeding in the pitchers, as well as changes in gene expression in a young species pair. We then examine signatures of natural selection on protein-coding sequences and gene expression levels in a comparative dataset encompassing the transcriptomes of 13 *Nepenthes* species, selected from a broad geographic range and disparate ecology (Figure 1). We show that thousands of genes experienced positive selection on amino acid substitutions, and shifted in their expression levels during the *Nepenthes* radiation. PFPs were preferentially targeted by positive selection and experienced shifts in expression levels, while genes upregulated during feeding were overrepresented among the genes showing expression changes between recently diverged species.

## Results

### The diversity of *Nepenthes* pitcher fluid proteins

We jointly analysed pitcher fluid proteomes and pitcher transcriptomes of twelve *Nepenthes* species. We found a much broader spectrum of PFPs across these species than has previously been known and that PFP composition and cumulative expression levels differ strongly among species (Figure 2). Transcriptome-predicted plant proteins found in pitcher fluids by mass spectrometry were clustered into high-order PFP "classes" with at least 40% amino acid sequence identity. We refer to proteins within these classes as isoforms. We inferred 27 PFP classes from 242 predicted protein sequences. Of these, twenty PFP classes have previously been identified, but seven classes are new, together with a large number of novel isoforms. The most diverse class are the Nepenthesins (Athauda et al. 1998, 2004), with a total of 19 isoforms across species and up to 16 isoforms per species (*N. khasiana*, *N. hemsleyana*), followed by the only recently discovered proline-cleaving Neprosins (Lee et al. 2016) with up to twelve, and BG3-like Glucanases with up to nine isoforms per species (Figure 2 left). Nepenthesins were hitherto understood as the most abundant and important digestive enzymes in *Nepenthes*, but we found that other PFPs occur in similar isoform diversity. Although Nepenthesins were the most highly expressed single PFP class (mRNA levels, Figure 2 right) in seven out of twelve species, they were always outnumbered by the cumulative transcript abundance of other PFP classes, except in *N. rafflesiana* t.f. The Nepenthesins were replaced as dominant PFP transcripts by Class IV Chitinases




| | alt. range [m] | soil | special prey | fluid viscosity | pitcher inside |
|---|---|---|---|---|---|
| 1 *N. pervillei* | 350-750 | granite rocks | | medium | waxy |
| 2 *N. khasiana* | 500-1500 | terrestrial | | low | waxy |
| 3 *N. gracilis* | 0-800 | sand, peat | | low | waxy |
| 4 *N. ampullaria* | 0-200(-2100) | sand, peat | leaf litter | low | not waxy |
| 5 *N. rafflesiana g.f.* | 0-300(-1000) | sand, peat | | medium | variable |
| 6 *N. rafflesiana t.f.* | 0-300 | sand, peat | | high | not waxy |
| 7 *N. hemsleyana* | 0-300 | sand, peat | bat guano | low | waxy |
| 8 *N. hamata* | 1800+ | moss, terrestrial | | low | waxy |
| 9 *N. jacqueliniae* | 1800+ | moss, epiphyte | | high | not waxy |
| 10 *N.* "*alata*" | unclear | terrestrial | | low | waxy |
| 11 *N. rajah* | 1500-2650 | ultramafic loam | faeces of tree shrews & rats | medium-low | not waxy |
| 12 *N. vogelii* | 1200-1500 | moss, epiphyte | | medium | not waxy |
| 13 *N. ephippiata* | 1800+ | moss, epiphyte | faeces of tree shrews | low | not waxy |

**Figure 1.** Top: Portraits of the characteristic pitcher traps produced by the 13 *Nepenthes* species examined in this study with a map of their provencance in Southeast Asia, India and the Seychelles. Bottom left: species tree estimated from 1,722 one-to-one orthologous genes for the 13 *Nepenthes* species rooted to the Venus flytrap, *Dionaea muscipula*. Node labels indicate Internode Certainty (IC, Kobert et al. 2016). Bottom right: major ecological and physiological traits. All *Nepenthes* prey on Arthropods, while some prefer special prey types. Data from Cheek & Jebb (2001) and own observations.

Left panel (pitcher fluid protein (PFP) classes):

| | total isoforms | *pervillei* | *khasiana* | *gracilis* | *ampullaria* | *rafflesiana* g.f. | *rafflesiana* t.f. | *hemsleyana* | *hamata* | *jacqueliniae* | *rajah* | *vogelii* | *ephippiata* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| total per species | | 30 | 56 | 7 | 10 | 32 | 22 | 75 | 11 | 7 | 56 | 37 | 32 |
| Nepenthesin | 19 | 8 | 16 | 1 | 1 | 9 | 11 | 16 | 3 | 5 | 12 | 11 | 7 |
| Neprosin | 16 | 8 | 9 | 2 | 1 | 8 | 4 | 12 | 0 | 2 | 8 | 10 | 4 |
| Glucan endo-1,3-beta-glucosidase BG3-like | 11 | 5 | 0 | 1 | 2 | 3 | 0 | 9 | 0 | 0 | 4 | 5 | 5 |
| Peroxidase | 8 | 1 | 3 | 0 | 0 | 1 | 0 | 7 | 2 | 0 | 2 | 1 | 0 |
| Thaumatin-like protein, cluster 2 | 6 | 4 | 2 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 1 |
| Chitinase Class IV | 5 | 1 | 4 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 4 | 0 | 0 |
| Purple acid phosphatase 27-like | 5 | 0 | 5 | 0 | 0 | 1 | 0 | 5 | 1 | 0 | 4 | 0 | 3 |
| Chitinase Class III | 4 | 1 | 2 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 2 | 0 | 2 |
| DOMON-domain protein | 4 | 1 | 1 | 0 | 1 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| Pathogenesis-related protein | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 0 |
| Purple acid phosphatase 20-like | 4 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 2 | 0 | 4 | 0 | 3 |
| S-like Ribonuclease | 4 | 1 | 3 | 0 | 0 | 1 | 2 | 1 | 1 | 0 | 3 | 3 | 1 |
| Serine carboxypeptidase 20-like | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| Serine carboxypeptidase type 3 | 3 | 0 | 3 | 0 | 0 | 1 | 0 | 3 | 0 | 0 | 2 | 1 | 1 |
| EG45/DPBB-like protein | 2 | 0 | 2 | 0 | 0 | 2 | 1 | 2 | 0 | 0 | 1 | 0 | 0 |
| Chitinase Class I | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Chitinase Class V | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| DENN-domain protein | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| GDSL esterase/lipase | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Glucan endo-1,3-beta-glucosidase 14-like | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Hevein-like | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| Nepenthesin-like | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Non-specifc lipid-transfer protein-like | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Pectinesterase | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| PPR-repeat protein | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| Serine carboxypeptidase 51-like | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Thaumatin-like protein, cluster 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

Right panel:

| | *pervillei* | *khasiana* | *gracilis* | *ampullaria* | *rafflesiana* g.f. | *rafflesiana* t.f. | *hemsleyana* | *hamata* | *jacqueliniae* | *rajah* | *vogelii* | *ephippiata* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| total per species | 2831 | 1133 | 1886 | 597 | 1031 | 1806 | 3197 | 1965 | 1398 | 3622 | 6654 | 2525 |
| Nepenthesin | 2.68 | 1.86 | 2.62 | 1 | 2.58 | 2.98 | 3.05 | 2.81 | 2.71 | 3.15 | 3.47 | 2.58 |
| Neprosin | 1.87 | 0.29 | 2.08 | 0.19 | 1.95 | 2.08 | 3 | 2.13 | 2.21 | 2.27 | 3.24 | 1.84 |
| Glucan endo-1,3-beta-glucosidase BG3-like | 2.02 | 1.01 | 1.77 | 1.33 | 1.26 | 0.97 | 1.66 | 0.69 | 0.9 | 1.64 | 2.4 | 1.37 |
| Peroxidase | 2.49 | 2.19 | 1.56 | 1.43 | 1.81 | 2.18 | 1.97 | 2.01 | 0.91 | 2.07 | 1.65 | 1.55 |
| Thaumatin-like protein, cluster 2 | 2 | 1.16 | 2.25 | 2.32 | 0.81 | 1.24 | 1.48 | 0.86 | 1.76 | 1.49 | 1.61 | 1.91 |
| Chitinase Class IV | 2.78 | 2.22 | 2.79 | 0.31 | 1.8 | 1.98 | 1.96 | 2.53 | 0 | 2.42 | 2.11 | 2.01 |
| Purple acid phosphatase 27-like | 1.97 | 0.41 | 0.93 | 0 | 0.89 | 1.27 | 1.82 | 1.22 | 0.37 | 1.65 | 1.86 | 0.52 |
| Chitinase Class III | 1.89 | 1.99 | 1.91 | 2.13 | 1.34 | 1.68 | 1.86 | 2.04 | 2.6 | 2.54 | 2.72 | 3.06 |
| DOMON-domain protein | 1.98 | 0 | 1.48 | 1.88 | 2.01 | 2.07 | 2.06 | 0 | 0 | 0 | 0 | 0 |
| Pathogenesis-related protein | 0 | 0.49 | 0 | 0.6 | 0.58 | 0.47 | 0.14 | 0.42 | 0 | 2.44 | 0.25 | 0.86 |
| Purple acid phosphatase 20-like | 2.5 | 1.24 | 1.96 | 0.98 | 1.85 | 1.83 | 1.42 | 1.63 | 1.07 | 2.5 | 2.36 | 2.53 |
| S-like Ribonuclease | 2.17 | 1.33 | 2.04 | 0.4 | 1.5 | 1.45 | 1.55 | 2.15 | 1.12 | 1.79 | 1.72 | 1.12 |
| Serine carboxypeptidase 20-like | 2.04 | 2.6 | 1.85 | 1.05 | 1.79 | 1.42 | 1.76 | 1.79 | 1.31 | 1.63 | 2 | 1.3 |
| Serine carboxypeptidase type 3 | 1.58 | 0.78 | 1.63 | 0 | 1.19 | 1.92 | 2.46 | 2.46 | 1.6 | 2.11 | 2.52 | 2.12 |
| EG45/DPBB-like protein | 1.27 | 0 | 0 | 1.72 | 0.32 | 0.27 | 1.83 | 0 | 0.94 | 1.51 | 1.42 | 0 |
| Chitinase Class I | 0.72 | 0.77 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 |
| Chitinase Class V | 0.73 | 0.97 | 0.49 | 0 | 0.83 | 0.23 | 0.3 | 0.7 | 0.48 | 0.64 | 0.99 | 1.57 |
| DENN-domain protein | 0.9 | 1.07 | 0.42 | 0.58 | 0.93 | 0.59 | 0.37 | 0.62 | 0.39 | 0.6 | 0.7 | 0.3 |
| GDSL esterase/lipase | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 |
| Glucan endo-1,3-beta-glucosidase 14-like | 1.79 | 2.05 | 0 | 1.38 | 1.28 | 1.28 | 1.65 | 1.6 | 1.55 | 1.54 | 1.78 | 1.59 |
| Hevein-like | 0 | 0 | 0 | 0 | 0 | 0.81 | 0 | 0 | 0 | 0 | 0 | 0 |
| Nepenthesin-like | 2.11 | 1.27 | 0.85 | 0.88 | 1.18 | 0.85 | 0.85 | 1.06 | 0.72 | 0.57 | 0.73 | 0.69 |
| Non-specifc lipid-transfer protein-like | 1.11 | 0 | 0.57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.99 |
| Pectinesterase | 1.78 | 0.96 | 0.99 | 0.72 | 1.54 | 1.5 | 1.53 | 0.98 | 2.07 | 2.4 | 1.88 | 1.95 |
| PPR-repeat protein | 1.11 | 1.19 | 0.69 | 0.61 | 1.08 | 1.17 | 1.15 | 1.18 | 0.86 | 1 | 0.77 | 0.65 |
| Serine carboxypeptidase 51-like | 0.45 | 0.53 | 1.02 | 0.94 | 1.13 | 0.73 | 1.13 | 0.71 | 0.56 | 0.99 | 1.23 | 0.58 |
| Thaumatin-like protein, cluster 1 | 0 | 0 | 0.4 | 0.58 | 0 | 0.52 | 0.45 | 0.76 | 0 | 0 | 0 | 0.27 |

**Figure 2.** Left panel: Plant-derived pitcher fluid protein (PFP) classes identified by mass spectrometry in pitcher fluids of 12 *Nepenthes* species seven days after elicitor treatment. Shown are numbers of distinct 'isoforms' per PFP class. Note that failure to detect proteins is not evidence of absence. Right panel: Gene expression levels for the same protein classes in the glandular pitcher walls three days after feeding with *Drosophila*, measured cumulatively over all homologs present in the transcriptomes of each species. Top row unit TMM-normalised TPM, below scaled by $\log_{10}(x+1)$. Trees above the panels indicate phylogenetic relationships among species.

in *N. pervillei* and *N. gracilis*, by Serine carboxypeptidase 20-like enzymes in *N. khasiana*, by Thaumatin-like proteins (cluster 2) in *N. ampullaria*, and by Class III Chitinases in *N. ephippiata* (Figure 2 right). The studied species displayed a more than eleven-fold variation in cumulative PFP expression across all classes and isoforms, exemplified by the low extreme *N. ampullaria*, a detritivorous species with unusually weak fluid acidity, and the insectivorous *N. vogelii* as the high extreme (Figure 2 right top row). Our results thus reveal hitherto unknown PFP diversity across the studied *Nepenthes* species.

In addition to the numerous isoforms in 20 previously known PFP classes, we here report seven novel PFP classes. A DENN-domain protein, detected in three species, was most similar to *Arabidopsis* SCD1 which may be involved in cytokinesis but also clathrin-mediated endocytosis and reaction to pathogens (McMichael et al. 2013). A PPR-repeat protein was detected in nine samples. Members of this large protein family are known to mediate organellar RNA editing (Barkan & Small 2014). A pectine methyl esterase with strongest similarity to *Arabidopsis* PME17 was found in *N. vogelii*. These enzymes are known to aide in pathogen defense and to modify the rheology of cell wall components (Sénéchal et al. 2014) and it is tempting to speculate that this enzyme is involved in

regulating the visco-elasticity of *Nepenthes* pitcher fluids (Gaume & Forterre 2007; Bonhomme et al. 2011). Four samples yielded a Nepenthesin-like protein with only 31% identity to any canonical Nepenthesin but with the same PFAM domain structure. A Class V Chitinase most similar to *A. thaliana* AT4G19810.1 was found in two species. Four samples contained Class I Chitinase proteins. These were already known to be expressed in pitcher tissues (Eilenberg et al. 2006), and underwent sub-functionalisation into carnivory versus pathogen defence in carnivorous Caryophyllales (Renner & Specht 2012), but the protein was previously not documented in *Nepenthes* digestive fluids. Two isoforms of an EG45/DPBB-like protein including signal peptides were found in six samples each. The identification of seven new PFP classes across diverse *Nepenthes* species reveals the complexity of plant carnivory. Variation among species in the number of PFP classes and isoforms clearly establishes that prey digestion is a highly variable quantitative trait in *Nepenthes*.

### Gene expression changes in feeding *Nepenthes* pitchers

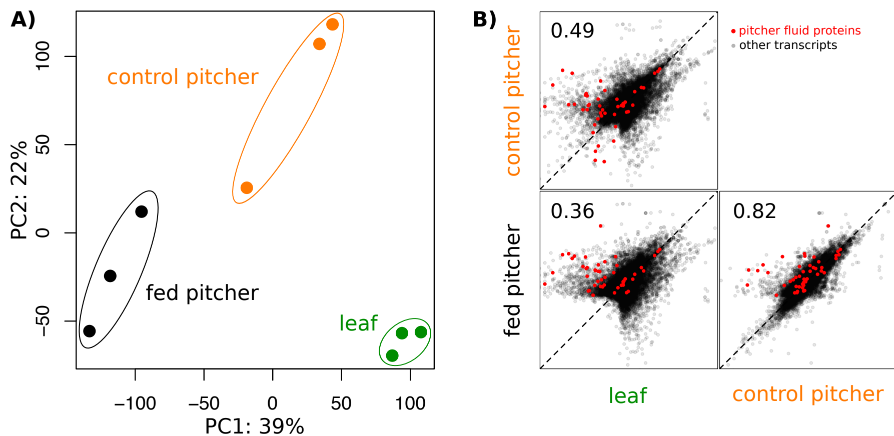



**Figure 3.** Comparison of gene expression based on 22,952 contigs between leaves, fed and control pitchers. A) Principal Component Analysis on normalised gene expression counts. B) Pairwise expression correlations between leaves, fed and control pitchers based on $\log_2$-scaled mean gene expression counts. Numbers are Pearson's correlation coefficients; all correlations with $p<2.2\times10^{-16}$; dotted lines indicate equal gene expression between compared tissues.

After prey has been degraded by PFPs, as well as by microbes and aquatic animals living in the pitcher fluid (e.g. Lam et al. 2017), *Nepenthes* pitchers absorb and assimilate the released molecules, similar to an animal's gut. Hence, numerous physiological processes in the pitcher tissue fulfil essential supportive tasks during

feeding. By comparing gene expression in fed pitchers against controls (unfed pichers) and against non-carnivorous leaves, we inferred candidate genes involved in these processes. Gene expression differed most strongly between fed pitchers and leaves, while control pitchers were intermediate along the first principal component (Figure 3A). As expected, PFPs typically showed higher transcript levels in fed than in control pitchers, and higher levels in pitchers than in leaves (Figure 3 B). These findings support the notion that PFP production is prey-induced (Buch et al. 2015; Yilamujiang et al. 2016). However, the observation that several PFPs appear more strongly expressed in leaves associated with fed pitchers than in the control pitchers may indicate that PFP expression between pitchers and associated leaves is not completely decoupled.

We found that 19.6% of the 20,182 contigs analysed were more strongly expressed in control than in fed pitchers. These are enriched for molecular functions (GO terms) related to photosynthesis, response to light stimulus, circadian clock, leaf development, and location in plastids and photosystems. In contrast, 14.7% of all contigs were more strongly expressed in fed pitchers than control pitchers and these were preferentially involved in peptide biosynthesis, translation, oxidoreduction, and ribosomal location. Besides the dominant up-regualtion of protein biosynthesis machinery, genes related to proteasomes, glycolysis, ATP-generation and cellular respiration were upregulated. The involvement of proteasomes could imply that part of the digestive process occurs intracellularly, acting on imported peptides or endocytotic vesicles, which have been documented in the pitcher glands (Adlassnig et al. 2012).

Gene expression differences between fed pitchers and leaves were much stronger than between fed and control pitchers but identified remarkably similar gene functions (Figure 3 A&B). Genes with higher transcript abundance in leaves (25.6% out of 20,595 contigs) were overall enriched for tasks in photosynthesis, carbon assimilation, chlorophyll synthesis, and chloroplastic locations. Transcripts with higher expression in pitchers (25.5%), on the other hand, were enriched for functions in aerobic respiration, nucleoside triphosphate metabolism, lipid oxidation, protein synthesis, and also for oxidoreductases and transmembrane-transporter ATPases, and preferentially located in mitochondria and ribosomes.

Our results reveal that inactive (unfed) pitchers may function similar to leaves and that feeding could be accomanied by a dramatic physiological shift from photosynthesis as the dominant process to proteins biosynthesis, proteolytic activity, and respiratory energy production, i.e. a more heterotrophic state. We hypothesise that *Nepenthes* mitigates the energetic costs of carnivory by switching the central task of its pitchers from photosynthetic carbon assimilation to carnivory in response to unpredictable prey capture. The ability to rapidly adjust pitcher physiology may represent a key adaptation that allowed *Nepenthes* species to reduce energetic costs associated with the pitcher traps.

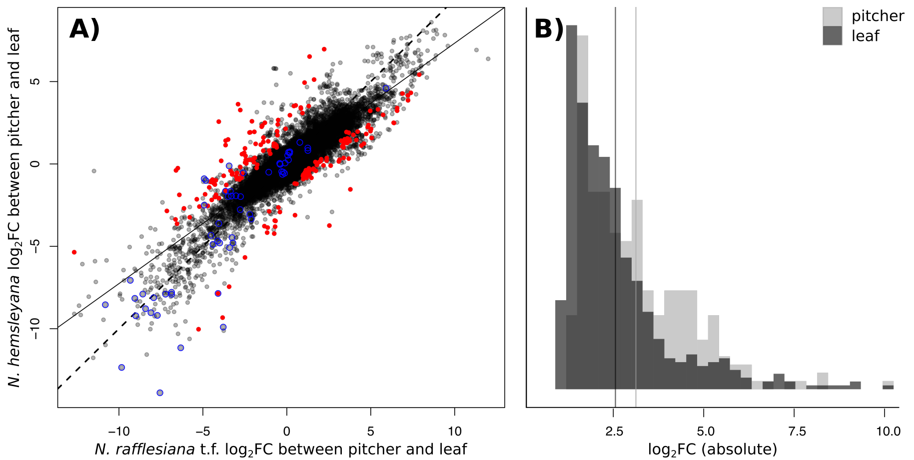



**Figure 4.** Fold-changes in expression levels between the fed pitchers and leaf tissues in the sister species *N. rafflesiana* t.f. and *N. hemsleyana*. A) correlation of fold-changes; red points: contigs showing significant tissue by species interaction, black points: interaction not significant; blue circles: PFPs; dotted line: 1:1 expectation; continuous line: linear model fit ($y = 0.73 \times x + 0.036$, $F_{1,\,25117} = 8.467\times10^{4}$, $p < 2.2\times10^{-16}$). B) Histograms of interspecific fold-changes ($\log_2$, absolute) for DEGs between pitchers (light grey) and leaves (dark grey); distribution means indicated by vertical bars.

### Gene expression differences between young sister species

We hypothesised that functional bias in genes differentially expressed between young sister species could be indicative of directional selection on regulatory mechanisms whereas neutral divergence should lead to functionally unrelated differentially expressed genes. We investigated divergence in gene expression between the incipient sympatric sister species *N. rafflesiana* t.f. and *N. hemsleyana* from two different perspectives. First, we fitted a two-way regression model with factors species and tissue (fed pitchers and leaves) to the gene expression counts of the two species. As may be expected, differences in expression between pitcher and leaf were highly correlated among *N. rafflesiana* t.f. and *N. hemsleyana* (Figure 4 A), and multi-dimensional scaling clustered tissues before species. However, 0.8% of 25,119 tested contigs showed significant species by tissue interaction effects, i.e. the strength or direction of fold-change between the tissues was species-specific (Figure 4 A red points). These contigs mark where the sister species diverged in their leaf vs. pitcher differentiation. Affected genes were functionally rather heterogeneous with the few enriched GO terms dominated by regulation of the actin cytoskeleton, gluconeogenesis, flavonoid biosynthesis, and oxidoreductase activity.

Second, we compared transcript levels between species for each of the two tissue types separately (common normalisation). Although fewer differentially

expressed genes (DEGs) were found between pitchers (1% of 25,119 contigs) than between leaves (2.1%), the $\log_2$ fold-change was on average 56% higher in comparisons of pitchers (mean fold-change pitchers = 8.75; mean fold-change leaves = 5.92; permutation-p = $2\times10^{-5}$; Figure 4 B). This was a robust result also observed among all expressed contigs (p = $2\times10^{-5}$). DEGs from the pitchers of the two species were enriched for functions in secondary metabolite biosynthesis, fatty acid transport, and developmental regulators, while DEGs from the leaves were enriched for the same functions, with the notable additions of steroid synthesis and chromoplast location. The latter may correspond to the deep red leaves of *N. hemsleyana* versus the bright green leaves of *N. rafflesiana* t.f. under identical conditions in the growth chamber.

Pitcher–leaf differentiation in gene expression was overall lower in *N. hemsleyana* than *N. rafflesiana* t.f., as indicated by the smaller than unity slope of the correlation (Figure 4 B). This pattern was apparently dominated by a more leaf-like transcriptome of *N. hemsleyana* pitchers rather than convergence of pitchers and leaves towards an intermediate transcriptome, as the sister species were still more similar for gene expression in leaves than in pitchers (Figure 4 B). Hence, the pitchers of *N. hemsleyana* are less divergent in expression from leaves than pitchers of *N. rafflesiana* t.f., which could reflect a weaker functional specialization for carnivory or reduced investment in digestion – an interesting possibility given that *N. hemsleyana* has 'outsourced' some of its prey digestion to small bats which roost in its pitchers and feed the plant with guano (Grafe et al. 2011; Schöner et al. 2017), while its sister species is an insectivore.

Interspecific divergence in transcriptome-wide gene expression was previously reported to vary among tissues in animals (Khaitovich et al. 2005; Gu & Su 2007; Brawand et al. 2011), potentially as a consequence of tissue-specific selection regimes or different proportions of tissue-specific versus universally expressed genes. Our results show that gene expression between *N. rafflesiana* t.f. and *N. hemsleyana* has diverged more strongly in pitchers than in leaves, which is consistent with stronger diversifying selection acting on pitchers or instead less intense stabilising selection allowing for more neutral divergence of the pitcher transcriptomes.

**Adaptive molecular evolution in *Nepenthes***

To understand whether adaptive evolution accompanied the diversification of *Nepenthes*, we built a comparative dataset with c. 25,000 clusters of homologous, expressed predicted proteins (irrespective of paralogy and orthology) from the de novo transcriptome assemblies of 13 species. We measured species-specific expression levels of these homolog clusters as the cumulative expression over all members of a homolog cluster (for 12 species, exlcuding *N. "alata"*). We hereafter refer to these homolog clusters as genes. We tested for positive selection on coding sequences, for shifts between two alternative gene expression level optima on internal branches of the phylogeny (Ornstein-Uhlenbeck process), and examined variation in expression levels based on the average of phylogenetically

**Table 1.** Interrelations between patterns of molecular evolution and the genetic basis of carnivory in *Nepenthes* (PFPs and carnivory-related genes), as tested in 2×2 contingency tables and permutation tests, respectively. Branch-site tests for positive selection on amino acid substitutions were conducted across 13 species. Gene expression levels were known for 12 species and tested for shifts in optimal expression using models of quantitative trait evolution (OU = Ornstein-Uhlenbeck process), as well as non-parametrically compared using phylogenetically independent contrasts (PICs; permutation tests on the difference in means). Speciation-related genes are defined here as genes with differential expression between *N. rafflesiana* t.f. and *N. hemsleyana*, in any of pitchers, leaves or with an interaction tissue × species (see previous section).

| | branch-site test | | OU shift of optimal expression level | | expression level mean of PICs | | speciation-related | |
|---|---|---|---|---|---|---|---|---|
| | sign. | n.s. | shift | no shift | mean | N | yes | no |
| PFPs | 15 | 34 | 14 | 35 | 4.8 | 49 | 5 | 24 |
| non-PFPs | 2673 | 22357 | 4113 | 21096 | 3.1 | 25209 | 253 | 8014 |
| | $\chi^2 = 18.3$<br>$p = 2\times10^{-5}$ | | $\chi^2 = 4.5$<br>$p = 0.03$ | | permutation test<br>$p = 2\times10^{-5}$ | | $\chi^2 = 14.9$<br>$p = 0.0001$ | |
| carnivory-related | 138 | 833 | 174 | 798 | 4.6 | 972 | 40 | 932 |
| not carnivory-related | 908 | 4620 | 886 | 4669 | 4.0 | 5555 | 112 | 5443 |
| | $\chi^2 = 2.8$<br>$p = 0.09$ | | $\chi^2 = 2.2$<br>$p = 0.14$ | | permutation test<br>$p = 2\times10^{-5}$ | | $\chi^2 = 15.1$<br>$p = 0.0001$ | |

independent constrasts (PICs). We found that PFPs and 'carnivory-related genes' (972 genes that were upregulated during feeding in the pitchers and also more strongly expressed in pitchers than in leaves) were over-represented among genes with signatures of adaptive molecular evolution.

Signatures of diversifying selection on amino acid substitutions (branch-site tests) were evident in 2,688 (10.7%) out of a total of 25,079 tested genes, when enforcing a strict branch length cut-off (0.1) to remove false-positives arising from alignment errors. These genes were mildly enriched for functions (GO terms) in immune system processes and programmed cell death. Shifts between two alternative gene expression level optima on internal branches of the phylogeny were detected for 4,127 (16.3%) out of 25,258 tested genes, with no enriched GO terms. We furthermore observed that adaptations at the coding sequence were slightly depleted among genes showing expression level shifts (1.5% out of 25,075 genes, expected 1.8%, $\chi^2_{df=1}$=8.5, $p<0.01$), indicating that positive selection in *Nepenthes* does typically not act jointly on regulatory and structural variation. Strong functional bias was evident among the genes in the top 5% of the transcriptome-wide distribution of gene expression variation (mean PICs): they are involved preferentially in translation, peptide synthesis, nitrogen metabolism, photosynthesis, and aerobic respiration, and localised in ribosomes, chloroplasts and mitochondria. Interestingly, these GO terms were very similar to those retrieved for gene expression changes induced by feeding.

We then tested our main hypothesis that genes underlying carnivory show accelerated molecular evolution in *Nepenthes*. We could match 49 genes of the comparative dataset to PFPs, 15 of which showed signatures of positive selection at the coding sequence level. This is an approximately threefold over-representation of this functional class among all genes exhibiting signatures of positive selection at the coding sequence (Table 1, first column). PFPs were furthermore over-represented among genes displaying shifts in expression level across the phylogeny, had elevated expression variation between species, and were over-represented among the genes diverging in expression in a young species pair (Table 1).

We identified as 'carnivory-related genes' those 972 genes with higher expression in fed pitchers compared to control pitchers or leaves in *N. rafflesiana* t.f. These genes were between species on average more variable in their expression (mean PICs) than other genes (Table 1), which may be a consequence of elevated phenotypic plasticity, relaxed stabilising selection, or diversifying selection. Interestingly, though, genes under positive selection or displaying shifts in expression were not overrepresented among carnivory-related genes (Table 1). We hypothesize that they encompass a high proportion of genes involved in fundamental cellular processes such as respiration and translation, and that these genes are under purifying selection. However, carnivory-related genes included three PFPs (a Nepenthesin, a purple acid phosphatase 20-like protein and a peroxidase) and were overrepresented among the genes that are differentially expressed between young sister species (Table 1), potentially indicating that prey digestion is a quantitative trait that quickly diverges between *Nepenthes* species with different prey spectra (insects versus bat guano).

## Discussion

Our combined proteomic and transcriptomic analysis of 13 *Nepenthes* species reveals that carnivory is a complex quantitative trait with substantial variation between species in both gene expression, as well as protein composition and abundance, and uncovers the signature of adaptive evolution in the coding sequences and expression levels of genes underlying carnivory. While numerous other genes not directly related to feeding and carnivory also exhibited adaptive divergence at the coding sequence or changed their expression levels during the *Nepenthes* radiation, those linked to feeding were in almost all our tests overrepresented – between incipient species as well as in the phylogenetic comparative dataset. These findings support the notion that plant carnivory underwent adaptive evolution following its origin and contributed to the adaptive divergence between species. Based on these results we propose that carnivory represent as physiological key adaptation trait in the radiation of *Nepenthes*.

Our study is among the first to test at the molecular level the hypothesis of the existence of a key adaptation trait in an evolutionary radiation. We base our argument for adaptive evolution on functional bias in genes showing elevated

sequence or expression level divergence rather than rates of change or speciation. Neutral diversification is not expected to affect functionally similar or interacting genes, but should instead result in elevated divergence of genes with unrelated functions. A recent study of heavy-metal tolerance in *Arabidopsis halleri* (He et al. 2016) concluded that the genetic basis of complex organism-level functions conferring ecological adaptation can indeed consistently display molecular signatures of positive selection, in this case in expression. Similarly, transcriptome-wide tests for convergent coding sequence adaptations discovered novel genes involved in $C_4$ photosynthesis in grasses (Huang et al. 2017), and comparative transcriptomics of the *Lupinus* radiation in the Americas revealed positive associations between phylogenetic diversification rate and coding sequence adaptations, as well as shifts in gene expression (Nevado et al. 2016). This study, however, could not directly test whether life history (perenniality), the suspected key adaptive trait in this radiation was indeed the target of diversifying selection during this radiation, because genes controlling this complex trait are largely unknown. To reliably infer the role of selection in driving adaptive divergence of key ecological or morphological traits thus still requires knowledge of the genes underlying a given trait.

In comparative studies of evolutionary radiations, tests of selection in both coding sequence and gene expression levels are still rarely applied on a genome-wide or transcriptome-wide scale (but see e.g. Roux et al. 2014). Our study exemplifies the strength of this approach if genes can be linked to the trait of interest. It allows discovering adaptations without bias from pre-conceived hypotheses by contrasting signatures of molecular evolution between different groups of genes in the same species, and may discover functional bias in genes that have experienced similar selection regimes (e.g. Warren et al. 2010). We expect that the reduced costs of sequencing, together with innovative statistical advances (e.g. Rohlfs et al. 2014; Rohlfs & Nielsen 2015), will stimulate wider applications of these pioneering comparative transcriptomic approaches (Brawand et al. 2011; Perry et al. 2012; Roux et al. 2015; Nevado et al. 2016). In contrast to many other studies focusing exclusively on signatures of selection in coding sequences, we also addressed gene expression evolution and found that greater expression change higher incidence of shifting expression optima strongly associated with a presumed adaptive trait. This finding supports the long-standing hypothesis that regulatory changes may drive adaptations (Romero et al. 2012).

We uncovered a strong transcriptomic reaction of *Nepenthes* traps to feeding. The trend away from photosynthesis towards proteolysis, protein synthesis and respiration is mirrored in the Venus flytrap (*Dionaea muscipula*, Bemm et al. 2016), the only other carnivorous plant in which the transcriptomics of feeding has been studied to date. Experiments are required to clarify to which extent these transcriptional changes translate to physiology. Of particular interest is the net energy balance of *Nepenthes* pitchers, and whether the increased respiratory activity during feeding sources metabolites from internal storage, normal photosynthetic

leaves, or from the prey to fuel their own digestion. The latter might resolve whether plant carnivory, thus far understood as a mineral acquisition strategy, involves partial heterotrophy. Recent experiments of gas-exchange and feeding with $^{13}C$-labelled prey indicate that *Dionaea* may indeed fuel its respiratory energy production with organic carbon from prey (Fasbender et al. 2017). Such studies give fascinating new insights into general carnivorous plant physiology, and could in the future include comparative experiments to understand how carnivorous plant physiology has adapted to different prey types and environments.

Finally, another highly important question posed by our results concerns the role of presumed adaptions in carnivory-related genes and PFPs in speciation. Are these candidate adaptive genes directly involved and possibly contributing to the evolution of reproductive isolation, e.g. by tapping new types of prey or better control of micro-organisms, or do they represent secondary adaptations that follow suit other major evolutionary changes? We speculate that both of these scenarios occurred among the 13 ecologically diverse *Nepenthes* studied here (Figure 1). Genes that exhibit adaptive evolution in comparative genome- or transcriptome-wide datasets at the level of evolutionary radiations should be tested for divergence between incipient species, as we did here, or at the level of populations. Such dual-level studies may reveal the overlap between the targets of positive selection over short and long evolutionary timescales, on both structural and regulatory variation, and improve our understanding of the predictability and repeatability of adaptive evolution.

## Methods

### Plants

Lowland species including those used for differential expression analyses (*N. ampullaria*, *N. hemsleyana*, *N. rafflesiana* t.f., *N. rafflesiana* g.f., *N. gracilis*) were cultivated in a growth chamber at a constant temperature of 27°C, 90-100% relative humidity, 12 hours illumination, and in *Sphagnum*-perlite substrate with tray watering. Plants were regularly fertilised and treated with pesticides but not for several weeks prior and during experiments. All other *Nepenthes* were grown under variable but generally cooler conditions, either in greenhouses under natural lights or in terraria under fluorescent bulbs.

Raw data for greenhouse-grown *N. "alata"* was previously published (Fukushima et al. 2017). These plants belong to a single clone originating in the Philippines, but no further details on the provenance are known (pers. comm. K. Fukushima, Sh. Yamada). According to current taxonomy (Cheek & Jebb 2013) this glabrous plant possibly represents *N. graciliflora* rather than *N. alata*.

### RNA treatment and sampling

For analysis of expression changes by feeding, we fed separate pitchers of three different genotypes of *N. rafflesiana* t.f. either with powdered and freeze-dried *Drosophila* flies (10 mg/ml of digestive fluid, "fed"), and different pitchers of the same individuals with demineralised water ("control" / "unfed"). These

treatmens were performed on mature pitchers (minimum 7d after opening of the lid) that were kept free of insects and semi-sterile by plugging the orifice with cotton whool upon opening. Treatments on the same plants were not performed simultaneously but separated by several weeks until new pitchers had grown, to avoid any possible systemic responses of feeding (Krausko et al. 2017). After 72 hours of exposure to the treatments, the submerged part of the glandular pitcher wall (directly in contact with digestive fluid) was harvested, the fluid discarded and tissue surfaces washed in 70% ethanol before flash-freezing in liquid nitrogen. The leaf tissue was a c. 4 $cm^2$ part of the leaf from which the "fed" pitchers grew, harvested simultaneously and identically to pitchers.

To test gene expression differences between young species, we used three different genotypes of *N. hemsleyana* (the sympatric sister species of *N. rafflesiana* t.f., see Chapter I of this thesis), on which we performed the "fed" treatment as above and also harvested their leaves. To form the comparative dataset, ten further species with a single sample each were likewise subject to the "fed" treatment as above.

Extensive testing unfolded that the best quality and quantity of total RNA from *Nepenthes* was obtained by the Total RNA Mini Kit (Plant) (Geneaid Biotech Ltd, New Taipei City, Taiwan) with the "PRB" lysis buffer. Tissues were ground to a fine powder in liquid nitrogen by mortar and pistil and subsequent (still frozen) milling with steel beads on a shaker mill (various). To increase RNA yield from a limited amount of tissue, lysation was conducted twice and the volumes united. A DNAse digest was performed after elution, and RNA extracts kept frozen until further use.

**RNA-seq libraries and sequencing**

Sequencing libraries were prepared from 0.5 µg total RNA using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA), according to the manufacturer's protocol. We sequenced two lanes with ten resp. 15 libraries each on the Illumia HiSeq 2500 for 126 bp paired-end reads at the Functional Genomics Center Zurich.

**Transcriptome assemblies**

In addition to the transcriptomes of 12 species sequenced for this study, we downloaded RNA-seq raw reads for *N. "alata"* (Fukushima et al. 2017) from NCBI Sequence Read Archive (seven 101bp paired-end Illumina libraries, accession numbers DRR051743-DRR051749). Transcriptomes were de novo assembled for each species separately using Trinity (Grabherr et al. 2011). The *Dionaea* transcriptome raw assembly v1.03 was downloaded from http://tbro.carnivorom.com (Bemm et al. 2016). Coding sequences and peptides were predicted from raw transcriptome assemblies with TransDecoder.LongOrfs v3.0.0 and TransDecoder.Predict (Grabherr et al. 2011). Predicted proteins were discarded if lacking similarity to known plant genes (blastp e-value $\leq 10^{-5}$; custom database of all proteins of all

Eudicotyledonae with sequenced genomes, NCBI Genbank accessed 6 June 2016, collapsed at 70% identity with CD-HIT, Li & Godzik 2006).

**Inference of homolog clusters and orthology**

We inferred homology and orthology of predicted *Nepenthes* and *Dionaea* genes following the tree-based pipeline of (Yang & Smith 2014) with slight modifications. In brief, DNA sequences were all-versus-all BLAST compared, and a table of reciprocal hit fractions (coverage percentage for either of two sequences) was constructed with cutoff value 0.5. Clusters were built using mcl (van Dongen 2001), with parameters -I 1.4 and e-value threshold $10^{-5}$. Clustered sequences were then aligned based on their amino acid sequence (MAFFT v7, Katoh & Standley 2013) and maximum-likelihood phylogenetic trees (Stamatakis 2014) estimated on back-translated DNA sequences. To establish finer sequence clusters, trees were pruned (maximum 0.1 expected substitutions) and split based on a maximum length for internal branches of 0.05 expected substitution, which clearly excludes any gene duplications older than the *Nepenthes* crown (at max. 30 million years, assuming a high substitution rate of $1.67\times10^{-8}$ per site per year). *Dionaea* was exempted from branch-length pruning because it consistently displayed much higher distances than within *Nepenthes*. Sequences from these clusters were submitted to the same procedure a second time, and the final clusters were considered homologs. Homologs with at most one sequence per taxon and at least five taxa were retained as 1-to-1 orthologs to build a species tree. These and additional clusters of homologs with at least four sequences, irrespective orthology or paralogy, were used to scan for selection at the coding sequence and expression levels (homolog clusters which we refer to as genes in the comparative dataset, in total c. 25,000 generated from *Nepenthes*). Inclusion of potential paralogs more than doubled the number of genes available for analysis, and is justified because adaptations are expected in both orthologs and paralogs (e.g. copy number variation, sub- and neo-functionalisation).

**Transcriptome and homolog cluster annotations**

Predicted proteins of the *N. rafflesiana* t.f transcriptome with similarity to plant genes and majority-consensus peptide sequences of 1-to-1 orthologs were compared to UniProtKB/TrEMBL (blastp, e-value $\leq 10^{-5}$, version 28 June 2016). TrEMBL is taxonomically much more diverse than UniProtKB/Swiss-Prot, which was insufficient for taxonomic classification of sequences, and computationally more feasible than NCBI nr. Proteins were discarded whose best TrEMBL hit was not to any genus of Embryophyta (NCBI Taxonomy, Entrez query "Embryophyta[subtree] AND genus[rank]", accessed 19 February 2017). This step identified large numbers of fungal and bacterial sequences, which were expected given non-sterile plant cultivation and the pitchers' microbiome blooming on *Drosophila* powder. Remaining proteins were compared to UniProtKB/Swiss-Prot (blastp, evalue $\leq 10^{-5}$, version 19 February 2017). GO terms for hits to Swiss-Prot were extracted from the UniProtKB/Swiss-Prot text database (uniprot_sprot.dat,

version 19 February 2017). Signal peptides and transmembrane domains were predicted with Phobius 1.01 (Käll et al. 2004) and SignalP 4.1 (Petersen et al. 2011).

The same procedure was applied to homolog clusters, represented by the longest sequence in the cluster.

**Species tree estimation**

We used 1-to-1 orthologous sequences from 13 *Nepenthes* and *Dionaea* to estimate a common species tree to be used in molecular evolution tests. There were 1,722 orthologs present in each of the 14 taxa, yielding a concatenated matrix of 2,480,430 DNA sites, with 99% occupancy. We estimated a maximum-likelihood tree on the partitioned alignment using RAxML v8.2.4 (Stamatakis 2014) with the GTRCAT substition model. The final tree was optimised and evaluated using the Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-like aLRT, RAxML -f J). Full support for all nodes was indicated. As a more informative indicator of phylogenetic conflict among loci, we calculated the "internode certainty" (IC), which for a given internode quantifies its frequency in a set of gene trees, jointly with the frequency of the most common conflicting bipartition (Kobert et al. 2016). To this end, separate gene trees were estimated for each partition (RAxML, GTRCAT) and IC annotated on the overall tree based on this collection of gene trees (RAxML -f i).

**Proteomics treatment and sampling**

The digestive process was induced in *Nepenthes* pitchers by known elicitors of the secretion of proteins and secondary metabolites (Eilenberg et al. 2006, 2010; Buch et al. 2015), and minerals known to be absorbed by the traps of *Nepenthes* (Schulze et al. 1999; Moran et al. 2010) or other carnivorous plants (Adamec 1997). Digestive fluids were supplemented with an elicitor solution, and contained at the beginning of experimental treatment 0.3 mg/ml colloidal chitin (product no. C9752, Sigma-Aldrich, Buchs, Switzerland; pulverised), 100 µM jasmonic acid (product no. 14631, Sigma-Aldrich, Buchs, Switzerland), 15 mM $NH_4NO_3$, 1.15 mM $KH_2PO_4$, 0.7 mM NaCl, 0.35 mM $MgSO_4$, and 0.35 mM $CaCl_2$. This mineral stoichiometry follows approximately that of typical terrestrial insects (Elser et al. 2000; Finke 2008), while the concentrations are based on the observation of up to 15 mM $NH_4^+$ during prey digestion in *N. bicalcarata* pitchers (Clarke & Kitching 1995), and presented no unusual osmotic stress as they remained well below typical total solute concentrations within plants cells (200-800 mM, Sitte et al. 1998).

At least seven days after opening and blocking pitcher orifices with cotton wool, digestive fluids in pitchers were topped up with $ddH_2O$ to a natural fluid level, volumes were measured using syringes, and mixed with 10% 10x elicitor stock solution. Syringe tips were guarded with soft silicone tubes to avoid tissue damage. Preliminary experiments showed that elicitor addition did not harm pitchers, but within 48 h caused a drop from pH 5-6 to pH 1-3, and a

slight increase in solute protein. Digestive fluids were harvested 7-8 days after treatment and sterile filtrated (Whatman syringe filter 0.2 µm, FP 30/0.2 CA-S, GE Healthcare), then stored at -20°C until further use. Before mass spectrometry, digestive fluids were evaporated under vacuum at room temperature, and 10-100 µg protein (Qubit Assay) were aliquoted and completely dried. The extreme visco-elasticity of some digestive fluids (e.g. *N. rafflesiana* t.f., *N. jacqueliniae*) was overcome by dissolving the dried residues in 30-50 µl 1M HCl and incubation for 7-15 min at 95°C (acid hydrolysis of polysaccharides).

**Protein identification by mass spectrometry**
Proteins in digestive fluids were identified using gel-free shotgun LC-MS/MS analysis in combination with *Nepenthes* transcriptome sequences as reference database. Dried or concentrated digestive fluids were dissolved in buffer (10 mM Tris, 2 mM CaCl2, pH 8.2), proteins were precipitated with TCA (10% final concentration), and the pellets were washed twice with cold acetone. Dry pellets were dissolved in 45 µl buffer (as above) and 5 µl trypsin (100 ng/µl in 10 mM HCl). Digestion was carried out either in a microwave instrument (Discover System, CEM) for 30 min at 5 W and 60°C or overnight at 37°C.

For LC-MS/MS analysis, samples were dissolved in 0.1% formic acid (Romil) and an aliquot of 5-10% was analysed on a nanoAcquity UPLC (Waters Inc.) connected to a Q Exactive mass spectrometer (Thermo Scientific) equipped with a Digital PicoView source (New Objective). Peptides were trapped on a Symmetry C18 trap column (5 µm, 180 µm x 20 mm, Waters Inc.) and separated on a BEH300 C18 column (1.7 µm, 75 µm x 150 m, Waters Inc.) at a flow rate of 250 nl/min using a gradient from 1% solvent B (0.1% formic acid in acetonitrile, Romil)/99% solvent A (0.1% formic acid in water, Romil) to 40% solvent B/60% solvent A within 90 min. Mass spectrometer settings were: Data dependent analysis, precursor scan range 350-1500 m/z, resolution 70k, maximum injection time 100 ms, threshold $3\times10^6$; fragment ion scan range 200-2000 m/z, resolution 35k, maximum injection time 120 ms, threshold $10^5$.

Tandem mass spectra were converted to Mascot generic format using proteowizard v.3.0.5759 with vendor peak picking option for MS/MS and deisotoped and deconvoluted using the H-Scorer script (Savitski et al. 2010), as automatised in FCC (Barkow-Oesterreicher et al. 2013). Mascot (Matrix Science, London, UK; version 2.5.1.3) was used to search peak data against a custom protein database assuming non-specific enzymatic digestion, because some *Nepenthes* fluid proteins can lysate others (Yang 2014b). Our database (2,803,176 entries) combined UniProtKB/Swiss-Prot with all (unfiltered) proteins predicted by TransDecoder from the transcriptomes of all twelve *Nepenthes* species we sequenced. Mascot was searched with a fragment ion mass tolerance of 0.03 Da, a parent ion tolerance of 10.0 PPM, and oxidation of methionine specified as a variable modification.

Scaffold (version 4.6.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95,0% probability by

the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 95,0% probability (Protein Prophet algorithm, Nesvizhskii et al. 2003) and contained at least one identified peptide. If multiple database proteins were not distinguished by MS/MS peptides, they were united as "protein groups" to satisfy parsimony.

**Annotation of pitcher fluid proteins**

Predicted proteins from transcriptomes that were detected in digestive fluids were scanned for PFAM domains using hmmer 3.1b1 (Eddy et al.) with an e-value threshold of $10^{-10}$. Signal peptides and transmembrane domains were predicted with SignalP v.4.1 (Petersen et al. 2011) and Phobius v.1.01 (Käll et al. 2004). Using BLAST (Camacho et al. 2009) and e-value thresholds of $10^{-5}$, sequences were compared to NCBI Genbank nr (version 31 March 2016), *Arabidopsis thaliana* proteins in UniProtKB (version 3 April 2016), and a custom database of previously reported *Nepenthes* digestive fluid proteins. This database of "reference pitcher fluid proteins" was established by searching for all nucleotide and protein sequences for *Nepenthes* on NCBI Genbank (accessed 10 October 2016), and exclusion of any orthogroups without evidence from digestive fluid proteomes by referring to the cited literature. Sequences from the more recent studies by Lee et al. (2016) and Fukushima et al. (2017) were also added. Sequences were flagged as "microbial" if the header or description of an NCBI Genbank nr hit did not contain any genus of Tracheophyta (downloaded from NCBI taxonomy, 25 March 2016).

For "protein groups", each sequence was separately annotated, but annotations were collapsed into a single representative "group annotation", if any of the following conditions were met: (1) any of the sequences were previously reported from *Nepenthes*, (2) all sequences contained the same PFAM domains, (3) all sequences matched the same *Arabidopsis* sequence, (4) all sequences had at least one PFAM domain in common. Otherwise, all of the dissimilar annotations were reported. Finally, a table was generated listing all identified protein groups with their annotations, and all analysed *Nepenthes* pitcher fluids, with their total unique peptide counts per protein group.

**Expression levels of homolog clusters**

We estimated expression levels for homolog clusters and PFP families across species by mapping sequence reads back to the species-specific denovo reference assembly, using BWA-MEM (Li 2013) with default settings. Non-primary (redundant) alignments were excluded using samtools (view -F 0x0100). Read counts were extracted with samtools idxstats and converted to tags per million (TPM, Li & Dewey 2011) by a custom R function, normalising for library size and transcript length. TPM were averaged for species with multiple replicates (*N. hemsleyana*, *N. rafflesiana* t.f.). Species lacking homologs in their transcriptome assembly were assigned the expression level zero, reflecting that transcript abundance was below a detectable level. TPM were then summed within samples over all members of

homolog clusters (see homology inference) to achieve a cumulative expression level. TPM values of PFP classes and homolog clusters were then normalised across samples using the TMM algorithm in edgeR (Robinson et al. 2010).

**Tests for diversifying selection on homolog cluster expression levels within the *Nepenthes* radiation (Ornstein-Uhlenbeck shifts)**

Gene expression level is a quantitative trait whose phylogenetic evolution can be modelled by a Brownian motion with the additional parameter of an 'optimum' value towards which the trait value is drawn, i.e. an Ornstein-Uhlenbeck (OU) process (Butler & King 2004). We tested for diversifying selection on each homolog clusters expression level in this phylogenetic comparative framework, similar to the approach by Nevado et al. (2016). Using the R package ouch (Butler & King 2004) and our *Nepenthes* species tree (Figure 1), we first fit an OU model with a single expression level optimum (representing stabilising selection). We subsequently fitted nine additional OU models with two expression level optima each which contained a single shift between optima placed at a single internal node of the species tree (diversifying selection, no backward-shifts). Fits of two-optima models were compared against the single-optimum fit using likelihood ratio tests. If after FDR correction (5%, Benjamini & Hochberg 1995) any of the two-optima fits was better than the single-optimum fit we concluded that the homolog cluster showed a signature of diversifying selection on its expression level. Terminal nodes (single species) were not tested for separate optima to reduce any possible effect of non-genetic variation (plasticity, technical artefacts) in the selection test.

**Phylogenetically independent contrasts (PICs) of homolog cluster expression levels**

As a parameter-free indicator of interspecific variation in gene expression level we chose the mean over the absolute phylogenetically independent contrasts (PICs) per gene (homolog cluster). For this TMM-normalised TPM read counts were used, and averaged over the replicates for *N. rafflesiana* t.f. and *N. hemsleyana*, or the single estimate for the remaining species. PICs (Felsenstein 1985) can be used to compare quantitative traits between species and are a method to remove possible non-independence of trait observations that might result from shared ancestry and neutral trait evolution (Brownian motion). PICs were calculated using the species tree (Figure 1) and the R package ape (Paradis et al. 2004), for TMM-normalised TPM read counts. Expression levels were averaged over the replicates for *N. rafflesiana* t.f. and *N. hemsleyana*, while we took the single estimate for the remaining non-replicated species.

**Differential gene expression tests**

Raw sequence reads of *N. rafflesiana* t.f. and *N. hemsleyana* were mapped against the raw, unfiltered Trinity transcriptome assembly for *N. rafflesiana* t.f. using BWA-MEM (Li 2013) with default settings. We did not quality-filter reads (Williams et al. 2016) or mappings, i.e. multiple placements of reads were tolerated, and

all taxonomic and technical ambiguities were retained. Mapped read counts per contig were extracted from .bam files with samtools idxstats, and here we removed contigs that did not contain predicted ORFs, or whose ORFs were not annotating with Embryophyte taxonomy in TrEMBL (see above).

We tested differential gene expression using edgeR (Robinson et al. 2010). The following comparisons were conducted: (1) fed–control pitchers in *N. rafflesiana* t.f., using three different genotypes each subjected to both treatments, tested in a glm with paired design; (2) pitcher–leaf comparison in *N. rafflesiana* t.f., samples and test design as before; (3) interspecific contrasts of the fed pitchers and leaves of *N. hemsleyana* and *N. rafflesiana* t.f., using first an interactive model (expression level $\sim$ species + tissue + species$\times$tissue), and then separate constrasts within each tissue. Contigs with cumulatively less than ten mappings per million and presence in less than three samples were excluded prior to testing, and counts were TMM-normalised. We used the tag-wise estimate of dispersion, and DGE was considered significant at FDR $\leq 0.05$ (Benjamini & Hochberg 1995).

**Tests for selection on coding sequence of homolog clusters**

We tested for the signal of positive selection in each gene (homolog cluster) based on the ratio of non-synonymous to synoymous substitutions in a codon alignment (dN/dS or $\omega$). Purifying (conservative) selection is expected to leave a signal of $\omega < 1$, neutral evolution is expected to produce $\omega = 1$, and positive (diversifying) selection can generate $\omega > 1$. We employed "branch-site" tests, which ask whether for a given branch in a phylogenetic tree, the codons (sites) of a given protein are better classified into only a single $\omega$-category or rather several different $\omega$-categories including a category of $\omega > 1$ (positive selection). However, we did not want to restrict our molecular evolution hypotheses to specific branches of *Nepenthes* gene trees. To get an overall picture, we conducted "branch-site" model tests for all branches in all of c. 25,000 gene trees and corrected for multiple testing (Anisimova & Yang 2007). We followed this multiple testing procedure in an optimised implementation called Adaptive Branch-Site Random Effects Likelihood (aBSREL) available in the HyPhy software package (Smith et al. 2015).

After exclusion of *Dionaea*, each homolog cluster within *Nepenthes* was aligned on the peptide sequence in MAFFT v7 (Katoh & Standley 2013) and the DNA alignment reconstituted. Alignments contained at least four sequences and 90% presence per codon. Because of the sensitivity of the branch-site test to alignment errors, we evaluated results by posterior filtering on the estimated length of branches yielding significant aBSREL tests, we rejected positive tests when the estimated branch length exceeded 0.1 expected substitutions – it is not conceivable that more substitutions could have accumulated during the relatively young *Nepenthes* crown. Furthermore, tests of molecular evolution can be sensitive to violations of the background species tree (Mendes & Hahn 2016), hence aBSREL results were conducted with background trees specific to each gene, i.e. estimated from their alignment.

### Testing interrelations between PFPs respectively carnviroy-related genes and signatures of selection

Gene categories were tested against the incidence of signatures of selection in the radiation of *Nepenthes* (branch-site tests, shifted expression level optima in Ornstein-Uhlenbeck models) using contingency tables ($\chi^2$-tests) as implemented in the function stats.chi2_contingency from the Scipy package (www.scipy.org). For the quantitative variable 'mean of absolute PICs of expression levels', we used permutation tests (R package perm, 100,000 permutations) to test the significance of difference between the means of distributions.

### Gene Ontology enrichment tests

Predicted transcriptome proteins were annotated with GO terms based on their nearest hit in Uniprot SwissProt (blastp, threshold evalue $\leq 10^{-5}$). We tested GO term overrepresentation in contigs of interest versus background contigs based on simple presence–absence using Fisher's exact test in the topGO package in R Bioconductor (Alexa et al. 2006). The "gene universe" was specific for each test and was exactly identical to the set of GO-annotated contigs within which contigs of interest were discovered (i.e. all contigs tested for DGE, or all branch-site tested orthologs). In one enrichment analysis –differentially expressed genes between fed and control pitchers – the superficially dominant gene functions in "fed" pitchers were all related to protein biosynthesis. We thus repeated enrichment tests after masking all contigs annotated with any GO term retrieved by the search for "translation", "ribosom" and "peptide synth" on http://geneontology.org/ [accessed on 3rd March 2017]. This strategy uncovered previously hidden but neverthless highly significantly enriched GO terms during feeding.

## Acknowledgements

We thank the Ministry of Industry and Primary Resources of Brunei Darussalam for permission to export plant material, and I. and S. Hartmeyer, M. Zehnder and U. Zimmermann for donation of plant material. Claudia Michel helped with the wet lab, and the Genetic Diversity Center Zürich provided computational support. The Functional Genomics Center Zurich is thanked for sequencing, and P. Hunziker for proteomics analyses.

## References

Adamec L (1997) Mineral nutrition of carnivorous plants: A review. *The Botanical Review*, **63**, 273-299.

Adlassnig W, Koller-Peroutka M, Bauer S et al. (2012) Endocytotic uptake of nutrients in carnivorous plants. *The Plant Journal*, **71**, 303-313.

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600-1607.

An CI, Fukusaki E-I, Kobayashi A (2001) Plasma-membrane $H^+$-ATPases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. *Planta*, **212**, 547-555.

An CI, Fukusaki E, Kobayashi A (2002) Aspartic proteinases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. *Planta*, **214**, 661-667.

Anisimova M, Yang Z (2007) Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution*, **24**, 1219-1228.

Athauda SBP, Inoue H, Iwamatsu A, Takahashi K (1998) Acid proteinase from *Nepenthes distillatoria* (Badura). In: *Aspartic Proteinases Advances in Experimental Medicine and Biology*. (ed James MNG), pp. 453-458. Springer US.

Athauda SBP, Matsumoto K, Rajapakshe S et al. (2004) Enzymic and structural characterization of Nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *Biochemical Journal*, **381**, 295-306.

Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology*, **65**, 415-442.

Barkow-Oesterreicher S, Türker C, Panse C (2013) FCC - An automated rule-based processing tool for life science data. *Source Code for Biology and Medicine*, **8**, 3.

Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. *Proceedings of the National Academy of Sciences*, **106**, 1133-1138.

Bemm F, Becker D, Larisch C et al. (2016) Venus flytrap carnivorous lifestyle builds on herbivore defense strategies. *Genome Research*, **26**, 812-825.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289-300.

Biteau F, Nisse E, Miguel S et al. (2013) A simple SDS-PAGE protein pattern from pitcher secretions as a new tool to distinguish *Nepenthes* species (Nepenthaceae). *American Journal of Botany*, **100**, 2478-2484.

Bonhomme V, Pelloux-Prayer H, Jousselin E et al. (2011) Slippery or sticky? Functional diversity in the trapping strategy of *Nepenthes* carnivorous plants. *New Phytologist*, **191**, 545-554.

Brawand D, Soumillon M, Necsulea A et al. (2011) The evolution of gene expression levels in mammalian organs. *Nature*, **478**, 343-348.

Brawand D, Wagner CE, Li YI et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375-381.

Buch F, Kaman WE, Bikker FJ, Yilamujiang A, Mithöfer A (2015) Nepenthesin protease activity indicates digestive fluid dynamics in carnivorous *Nepenthes* plants. *PLOS ONE*, **10**, e0118853.

Buch F, Rott M, Rottloff S et al. (2013) Secreted pitfall-trap fluid of carnivorous *Nepenthes* plants is unsuitable for microbial growth. *Annals of Botany*, **111**, 375-383.

Butler MA, King AA (2004) Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, **164**, 683-695.

Camacho C, Coulouris G, Avagyan V et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cheek MR, Jebb MHP (2001) Nepenthaceae. *Flora Malesiana*, **15**, 1-157.

Cheek M, Jebb M (2013) Typification and redelimitation of *Nepenthes alata* with notes on the *N. alata* group, and *N. negros* sp. nov. from the Philippines. *Nordic Journal of Botany*, **31**, 616-622.

Clarke CM, Kitching RL (1995) Swimming ants and pitcher plants - a unique ant-plant interaction from Borneo. *Journal of Tropical Ecology*, **11**, 589-602.

Darwin C (1875) *Insectivorous plants*. J. Murray, London.

van Dongen SM (2001) *Graph clustering by flow simulation*. PhD Thesis, Utrecht University, The Netherlands.

Eddy SR, Wheeler TJ, the HMMER development team *HMMER*. http://hmmer.org/.

Eilenberg H, Pnini-Cohen S, Rahamim Y et al. (2010) Induced production of antifungal naphthoquinones in the pitchers of the carnivorous plant *Nepenthes khasiana*. *Journal of Experimental Botany*, **61**, 911-922.

Eilenberg H, Pnini-Cohen S, Schuster S, Movtchan A, Zilberstein A (2006) Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. *Journal of Experimental Botany*, **57**, 2775-2784.

Ellison AM, Gotelli NJ (2009) Energetics and the evolution of carnivorous plants – Darwin's 'most wonderful plants in the world.' *Journal of Experimental Botany*, **60**, 19-42.

Elser JJ, Fagan WF, Denno RF et al. (2000) Nutritional constraints in terrestrial and freshwater food webs. *Nature*, **408**, 578-580.

Enard D, Messer PW, Petrov DA (2014) Genome-wide signals of positive selection in human evolution. *Genome Research*, **24**, 885-895.

Fasbender L, Maurer D, Kreuzwieser J et al. (2017) The carnivorous Venus flytrap uses prey?derived amino acid carbon to fuel respiration. *New Phytologist*, **214**, 597-606.

Felsenstein J (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1-15.

Finke MD (2008) Nutrient Content of Insects. In: *Encyclopedia of Entomology* (ed Capinera, JL), p. . Springer, Dordrecht, The Netherlands.

Fraser HB (2011) Genome?wide approaches to the study of adaptive gene expression evolution. *BioEssays*, **33**, 469-477.

Frazier CK (2000) The enduring controversies concerning the process of protein digestion in *Nepenthes* (Nepenthaceae). *Carnivorous Plant Newsletter*, **29**, 56-62.

Fukushima K, Fang X, Alvarez-Ponce D et al. (2017) Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nature Ecology & Evolution*, **1**, 0059.

Gaume L, Forterre Y (2007) A viscoelastic deadly fluid in carnivorous pitcher plants. *PLOS ONE*, **2**, e1185.

Givnish TJ (2015) Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, **207**, 297-303.

Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-652.

Grafe TU, Schöner CR, Kerth G, Junaidi A, Schöner MG (2011) A novel resource-service mutualism between bats and pitcher plants. *Biology Letters*, **7**, 436-439.

Gu X, Su Z (2007) Tissue-driven hypothesis of genomic evolution and sequence-expression correlations. *Proceedings of the National Academy of Sciences*, **104**, 2779-2784.

Hatano N, Hamada T (2008) Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteome Research*, **7**, 809-816.

Hatano N, Hamada T (2012) Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteomics*, **75**, 4844-4852.

He F, Arce AL, Schmitz G et al. (2016) The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the *Arabidopsis* genus. *Molecular Biology and Evolution*, **33**, 2088-2101.

Hodges SA, Derieg NJ (2009) Adaptive radiations: From field to genomic studies. *Proceedings of the National Academy of Sciences*, **106**, 9947-9954.

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995-1016.

Hooker JD (1874) The carnivorous habits of plants. *Nature*, **10**, 366-372.

Huang P, Studer AJ, Schnable JC, Kellogg EA, Brutnell TP (2017) Cross species selection scans identify components of $C_4$ photosynthesis in the grasses. *Journal of Experimental Botany*, **68**, 127-135.

Juniper BE, Robins RJ, Joel DM (1989) *The carnivorous plants*. Academic Press, London.

Käfer J, Mousset S (2014) Standard sister clade comparison fails when testing derived character states. *Systematic Biology*, **63**, 601-609.

Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, **338**, 1027-1036.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772-780.

Khaitovich P, Hellmann I, Enard W et al. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, **309**, 1850-1854.

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107-116.

Kobert K, Salichos L, Rokas A, Stamatakis A (2016) Computing the internode certainty and related measures from partial gene trees. *Molecular Biology and Evolution*, **33**, 1606-1617.

Krausko M, Perutka Z, Šebela M et al. (2017) The role of electrical and jasmonate signalling in the recognition of captured prey in the carnivorous sundew plant *Drosera capensis*. *New Phytologist*, **213**, 1818-1835.

Krol E, Płachno BJ, Adamec L et al. (2012) Quite a few reasons for calling carnivores "the most wonderful plants in the world". *Annals of Botany*, **109**, 47-64.

Lam WN, Chong KY, Anand GS, Tan HTW (2017) Dipteran larvae and microbes facilitate nutrient sequestration in the *Nepenthes gracilis* pitcher plant host. *Biology Letters*, **13**, 20160928.

Lan T, Renner T, Ibarra-Laclette E et al. (2017) Long-read sequencing uncovers the adaptive topography of a carnivorous plant genome. *Proceedings of the National Academy of Sciences*, 201702072.

Lee L, Zhang Y, Ozar B, Sensen CW, Schriemer DC (2016) Carnivorous nutrition in pitcher plants (*Nepenthes* spp.) via an unusual complement of endogenous enzymes. *Journal of Proteome Research*, **15**, 3108-3117.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 1303.3997 [q-bio].

Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.

Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

Losos JB, Mahler DL (2010) Adaptive radiation: The interaction of ecological opportunity, adaptation, and speciation. In: *Evolution Since Darwin: The First 150 Years.* (eds Bell MA, Futuyama DJ, Eanes WF, Levinton JS), pp. 381-420. Sinauer Assoc., Sunderland, MA, USA.

McMichael CM, Reynolds GD, Koch LM et al. (2013) Mediation of clathrin-dependent trafficking during cytokinesis and cell expansion by *Arabidopsis* STOMATAL CYTOKINESIS DEFECTIVE proteins. *The Plant Cell*, **25**, 3910-3925.

McPherson S (2009) *Pitcher plants of the Old World*. Redfern Natural History, Poole.

Mendes FK, Hahn MW (2016) Gene tree discordance causes apparent substitution rate variation. *Systematic Biology*, **65**, 711-721.

Moran JA, Clarke CM (2010) The carnivorous syndrome in *Nepenthes* pitcher plants. *Plant Signaling & Behavior*, **5**, 644-648.

Moran JA, Hawkins BJ, Gowen BE, Robbins SL (2010) Ion fluxes across the pitcher walls of three Bornean *Nepenthes* pitcher plant species: flux rates and gland distribution patterns reflect nitrogen sequestration strategies. *Journal of Experimental Botany*, **61**, 1365-1374.

Naranjo S, Smith JD, Artieri CG et al. (2015) Dissecting the genetic basis of a complex cis-regulatory adaptation. *PLOS Genetics*, **11**, e1005751.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry*, **75**, 4646-4658.

Nevado B, Atchison GW, Hughes CE, Filatov DA (2016) Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nature Communications*, **7**, 12384.

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197-218.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.

Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology*, **14**, e1002379.

Pereira CG, Almenara DP, Winter CE et al. (2012) Underground leaves of *Philcoxia* trap and digest nematodes. *Proceedings of the National Academy of Sciences*, **109**, 1154-1158.

Perry GH, Melsted P, Marioni JC et al. (2012) Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research*, **22**, 602-610.

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**, 785-786.

Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS ONE*, **9**, e89543.

Renner T, Specht CD (2012) Molecular and functional evolution of class I chitinases for plant carnivory in the Caryophyllales. *Molecular Biology and Evolution*, **29**, 2971-2985.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139-140.

Rohlfs RV, Harrigan P, Nielsen R (2014) Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Molecular Biology and Evolution*, **31**, 201-211.

Rohlfs RV, Nielsen R (2015) Phylogenetic ANOVA: The expression variance and evolution model for quantitative trait evolution. *Systematic Biology*, **64**, 695-708.

Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, **13**, 505-516.

Rottloff S, Miguel S, Biteau F et al. (2016) Proteome analysis of digestive fluids in *Nepenthes* pitchers. *Annals of Botany*, **117**, 479-495.

Rottloff S, Stieber R, Maischak H et al. (2011) Functional characterization of a class III acid endochitinase from the traps of the carnivorous pitcher plant genus, *Nepenthes*. *Journal of Experimental Botany*, **62**, 4639-4647.

Roux J, Privman E, Moretti S et al. (2014) Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, **31**, 1661-1685.

Roux J, Rosikiewicz M, Robinson?Rechavi M (2015) What to compare and how: Comparative transcriptomics for evo?devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **324**, 372-382.

Sanderson MJ, Donoghue MJ (1996) Reconstructing shifts in diversification rates on phylogenetic trees. *Trends in Ecology & Evolution*, **11**, 15-20.

Savitski MM, Mathieson T, Becher I, Bantscheff M (2010) H-score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. *Journal of Proteome Research*, **9**, 5511-5516.

Schöner CR, Schöner MG, Grafe TU et al. (2017) Ecological outsourcing: A pitcher plant benefits from transferring pre?digestion of prey to a bat mutualist. *Journal of Ecology*, **105**, 400-411.

Schulze W, Frommer WB, Ward JM (1999) Transporters for ammonium, amino acids and peptides are expressed in pitchers of the carnivorous plant *Nepenthes*. *The Plant Journal*, **17**, 637-646.

Sénéchal F, Wattier C, Rustérucci C, Pelloux J (2014) Homogalacturonan-modifying enzymes: structure, expression, and roles in plants. *Journal of Experimental Botany*, **65**, 5125-5160.

Sitte P, Strasburger E, Ziegler H, Ehrendorfer F, Bresinsky A (1998) *Strasburger - Lehrbuch der Botanik für Hochschulen*. Spektrum Akademischer Verlag, Heidelberg.

Smith MD, Wertheim JO, Weaver S et al. (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, **32**, 1342-1353.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312-1313.

Takahashi K, Tanji M, Shibata C (2007) Variations in the content and isozymic composition of Nepenthesin in the pitcher fluids among *Nepenthes* species. *Carnivorous Plants Newsletter*, **36**, 73-76.

Takeuchi Y, Salcher MM, Ushio M et al. (2011) In situ enzyme activity in the dissolved and particulate fraction of the fluid from four pitcher plant species of the genus *Nepenthes*. *PLOS ONE*, **6**, e25144.

Voelckel C, Gruenheit N, Lockhart P (2017) Evolutionary transcriptomics and proteomics: insight into plant adaptation. *Trends in Plant Science*, **22**, 462-471.

Warren AS, Anandakrishnan R, Zhang L (2010) Functional bias in molecular evolution rate of *Arabidopsis thaliana*. *BMC Evolutionary Biology*, **10**, 125.

Whittall JB, Hodges SA (2007) Pollinator shifts drive increasingly long nectar spurs in columbine flowers. *Nature*, **447**, 706-709.

Williams CR, Baccarella A, Parrish JZ, Kim CC (2016) Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*, **17**, 103.

Yang Z (2014a) *Molecular Evolution: A Statistical Approach*. Oxford University Press, Oxford, United Kingdom?; New York, NY, United States of America.

Yang M (2014b) *Towards a biochemical reconstitution of Nepenthes pitcher fluid for the treatment of celiac disease.* PhD Thesis. University of Calgary, Calgary, Alberta, Canada.

Yang Y, Smith SA (2014) Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, **31**, 3081-3092.

Yilamujiang A, Reichelt M, Mithöfer A (2016) Slow food: insect prey and chitin induce phytohormone accumulation and gene expression in carnivorous *Nepenthes* plants. *Annals of Botany*, **118**, 369-375.

## Author Contributions

MS conducted the experiments, the bioinformatics and statistical analyses, and some of the wet lab procedures; FM and TUG contributed logistic support and materials; MS and AW conceived and designed the study, and wrote the manuscript.

# General Discussion

Whether the origin of species is a consequence of stochastic processes (e.g. Mayr 1942), an argument recently revived because of the clock-like accumulation of deep phylogenetic lineages (Hedges et al. 2015), or is instead driven by ecological and physiological adaptations, as proposed by Darwin (1859), remains a contentious debate in evolutionary biology. Chance arguably plays an important role as the source of spontaneous mutations, but adaptations are ubiquitous. The evolution of reproductive isolation, a key step in the speciation process, might be de-coupled from the evolution of adaptations and may accrue independently after speciation is completed (e.g. Nosil 2012). Alternatively, reproductive isolation may evolve together with adaptations that contribute to divergence and ultimately speciation. A major challenge is the empirical identification of important phenotypic and molecular targets of selection during evolutionary diversification. Plant carnivory is an enigmatic adaptive trait whose evolution is still little understood, in particular whether further adaptive evolution diversified this trait subsequent to its origin, and whether carnivory is associated with speciation.

To disentangle the roles of adaptive and neutral processes in the evolution of species and functional traits, I chose the *Nepenthes* pitcher plants as study system because they form species-rich communities that may stably co-exist in the same habitat despite incomplete reproductive isolation. Although this situation may be common in plants and other organisms that undergo rapid evolutionary diversification, studies of such complex multi-species scenarios are underrepresented in the literature, in favour of studies on pairs of incipient species that segregate along environmental gradients or occupy different ecological niches. Carnivory in *Nepenthes* is a fascinating system to study functional traits at the sequence level, because central physiological functions in carnivory are based on relatively simple molecular processes (e.g. the production and secretion of digestive enzymes), and it is thus possible to relate genes to essential aspects of this complex phenotype. In addition, sex determination in *Nepenthes* offers an opportunity to study an independent origin of dioecy that has not led to sex chromosome heteromorphy, which could shed new light on conceptual issues related to sex chromosome evolution (e.g. Charlesworth 2016).

Prior to this project, it was noted that *Nepenthes* species frequently occur in mixed stands of two or more morphospecies with occasional natural hybridisation. Functional differences in the pitcher traps, which might affect plant fitness, were reported between some of these coexisting *Nepenthes* species. Here I asked how species could retain their separate phenotypes despite an apparent potential to hybridise and back-cross. In chapter I of this thesis I demonstrated that *Nepenthes* morphospecies that coexist in ecological communities are genetically distinct, that

they span a large range of divergence levels, and that natural hybridisation indeed occurs. Importantly, I found evidence that hybridisation is not only associated with recent, anthropogenic habitat disturbance, but instead that introgression has occurred over much longer evolutionary time scales and among multiple species. While researchers have speculated that introgression between coexisting *Nepenthes* species might occur and make them 'less specialized' (Peng & Clarke 2015), the full evolutionary implications of introgression, if occurring extensively and long-term as suggested in this thesis, were previously not realized. The persistence of *Nepenthes* species in the face of gene flow, and of differences in trap morphology between some of them, can now be interpreted as the result of divergent natural selection. Together, my results suggest the existence of adaptive diversification processes in local *Nepenthes* communities that are linked to the carnivorous syndrome. It is thus possible that plant carnivory is not only adaptive in itself, but also contributing to adaptive diversification of the *Nepenthes* lineage.

This thesis suggests the occurrence of ongoing speciation processes among multiple *Nepenthes* species and should be followed in the future by experimental as well as more detailed genomic studies to elucidate the driving factors and consequences. Further work is required to understand which genetic, ecological or physiological mechanisms contribute to reproductive barriers that ultimately maintain species boundaries in the face of gene flow. Only a comprehensive assessment of intrinsic and ecological components of reproductive isolation in *Nepenthes* can test whether carnivory contributes to reduced hybrid fitness, and how relevant this barrier is. The strength of pre-zygotic (pollination, fertilisation) and early post-zygotic isolation barriers (e.g. hybrid seed failure: Oneal et al. 2016; Florez-Rueda et al. 2016; hybrid necrosis: Bomblies & Weigel 2007) could be estimated in experimental crosses and the incidence of hybrid offspring in wild collected seeds. Ecological post-zygotic reproductive barriers could be quantified in transplant experiments, by measuring growth and survival of hybrids relative to parental species in intact and disturbed habitats. The logistic challenges of such studies may be overcome with the help of local field assistants, and vegetative propagation of wild-collected species and hybrids rather than starting from seeds.

Furthermore, the pitcher traps of *Nepenthes* need to be characterised from developmental, geometric and physiological perspectives to understand how carnivory, especially functional morphology, is constrained by physical and energetic factors (Díaz et al. 2016). Cost-benefit leaf economy models (e.g. Reich et al. 2003) could then be used to understand whether different morphologies may indeed represent alternative optima of efficiency in carnivory.

Whole-genome sequences of *Nepenthes* would facilitate development of a detailed picture of the genomic landscape of divergence with gene flow. Latest long-read sequencing technology should now deliver reasonable results even for the relatively large *Nepenthes* genomes (c. 650 - 1,500 Mb for the most relevant species studied here, Knápek 2012). This resource would allow localisation of regions of gene flow and resistance in multiple species. In particular, genome

sequences would greatly improve a genomic outlier study, to clarify whether multiple species pairs under divergence with gene flow (DGF) share outlier regions ("genomic speciation islands"), or rather base their species identity on different loci with each interacting DGF partner. Ideally, genome assemblies could be combined with linkage maps and an analysis of quantitative trait loci (QTLs) for candidate adaptive traits.

This thesis also demonstrated the feasibility to uncover neutral and adaptive population genetic patterns in *Nepenthes*. Such information may help to improve taxonomy and conservation of the genus in the future.

In the third chapter of this thesis, I explored carnivory in *Nepenthes* from a combined transcriptomic and proteomic perspective and asked whether this trait had been targeted by positive selection during the radiation of the genus. This project is a contribution to the still small number of studies scanning whole genomes or transcriptomes for molecular signatures of adaptation during evolutionary radiations (Roux et al. 2014; Brawand et al. 2014; Pease et al. 2016; Nevado et al. 2016). Previous studies have demonstrated that large proportions of proteins in a genome can be targets of positive selection at both structural and regulatory levels during rapid radiations, and that there may be collective functional biases among these genes. However, the present study is different in that it focuses on the genetic basis of a suspected key adaptation trait. Because little was previously known about the genes whose products take on the physiological tasks of feeding in carnivorous plants, experiments were conducted together with transcriptomics and proteomics to identify relevant proteins and their underlying genes. The molecular functions of genes that changed in their expression upon feeding point to profound physiological changes induced by feeding. But the presumably most interesting insight from this study was that genes involved in prey digestion are overrepresented among the many thousands of genes under positive selection in the *Nepenthes* radiation, most of which is considerably younger than the crown age of the genus (17-24 My, see Supplementary Material to chapter II). We thus posit that genome-wide or transcriptome-wide analyses can identify key adaptive traits in evolutionary radiations, if their genetic basis is known.

The comparative dataset does not allow to infer whether phases of positive selection on genes with functions in carnivory occurred during early stages of speciation, possibly promoting reproductive isolation, or subsequent to the establishment of reproductive isolation. The former scenario is, however, quite possible, as regulatory differences between recently diverged species were also enriched for genes underlying the proposed key trait. Hence, two distinct aspects of the carnivorous syndrome – prey digestion and trap morphology – could be among the traits with a role in the ongoing speciation processes uncovered in chapter I of this thesis.

More pitcher fluid proteins were identified here than in previous studies, presumably for two main reasons. First, we used transcriptomes of 12 *Nepenthes* species as the reference database for protein identification, while others relied

on unspecific protein databases or single transcriptomes. Future studies could re-analyse existing mass spectrometry data (Rottloff et al. 2016). A second improvement was the omission of proteome fractionation by gel electrophoresis prior to mass-spectrometry – semi-sterile *Nepenthes* digestive fluid induced with a special elicitor contained at most a few hundred different proteins. However, the long-chain polysaccharides in some species' fluids were problematic and impeded even better protein discovery. Future studies should experimentally assess how the inferred adaptive changes in amino acid and expression levels affect physiology and ecology, highlighting the need for comparative experiments (see Barrett & Hoekstra 2011). Functional characterisation of pitcher fluid proteins is possible by heterologous expression in bacteria (e.g. Schraeder et al. 2017). These and also non-secretory carnivory-related genes could furthermore be expressed in model plants (e.g. *Arabidopsis*) to elucidate their general plant physiological properties.

*Nepenthes* are the best studied carnivorous plants in terms of digestive and antimicrobial proteins (Athauda et al. 1998, 2004; Eilenberg et al. 2006; Hatano & Hamada 2008, 2012; Rottloff et al. 2016; Lee et al. 2016; Fukushima et al. 2017), but many more could be present in other carnivorous plant genera (Płachno et al. 2006; Adlassnig et al. 2011), which have not yet been studied in detail (but see Fukushima et al. 2017). The combination of protein mass-spectrometry with transcriptomics, as demonstrated here, could improve this situation in the near future.

A second focus during this thesis was put on the evolution of dioecy and plant sex chromosomes. Comparative studies of sex allocation, sex determination and sex chromosome evolution, aiming to test general hypotheses about factors driving the evolution of dioecy and its consequences in plants, are limited by the current scarcity of evolutionarily independent study systems. Information about dioecy in *Nepenthes* prior to this study was essentially limited to reports of inflorescence differences between the sexes and indications that karyotypes did not reveal heteromorphic sex chromosomes (Heubl & Wistuba 1997) despite stable dioecy (Clarke 2001).

This project uncovered a genetic basis of dioecy in *Nepenthes* involving X and Y sex chromosomes, and identified several sex-linked, expressed genes. Consistent results were obtained in three species including the phylogenetically most basal one, indicating that the presumed XY system is as old as the crown age of the genus, which was newly dated with a molecular-clock approach. In the course of this work, a highly efficient and generally applicable statistical method to extract sex-specific loci from noisy next generation sequencing data was developed, which amends traditional approaches for sex chromosome discovery (segregation analysis in families, cytology) by genome-scans for sex-specificity in natural populations. This method, in combination with RAD-seq or other high-throughput methods such as RNA-seq and whole genome sequencing, may improve the development of markers for molecular sexing used in research, conservation and agriculture.

Knowing that males are the heterogametic sex and with reliable Y-chromosomal markers at hand, future studies may use long-read sequencing technology and diploid genome assemblers (e.g. Chin et al. 2016) on male *Nepenthes*, to fully characterise X and Y chromosome divergence. This may provide insights into why these relatively old plant sex chromosomes have not evolved large-scale heteromorphy, and provide further candidates for plant sex determining genes.

Crucial for further understanding of dioecy in *Nepenthes* will be an investigation into sexual dimorphism (Barrett & Hough 2013), starting with an account of physiological and ecological sexual dimorphism and the costs of male and female reproduction, which could indicate sexual conflict and sexual selection (e.g. Delph et al. 2002; Moore & Pannell 2011). *Nepenthes* species differ in their level of sexual dimorphism, including for example inflorescence size and flower number, but these patterns require more thorough quantification. Such differences may uncover different levels of sexual selection. In *Silene latifolia*, sexual dimorphism evolves at the level of local populations (Yu et al. 2011). However, *Silene* sex chromosomes are much larger and hence less accessible for genomics. A comparative approach using X and Y sequences of several *Nepenthes* species could be used to test whether in plants the strength of sexual dimorphism, a surrogate of past sexual selection strength, correlates with the size or gene content of the sex chromosomes, as recently suggested for guppies (Wright et al. 2017).

Another question that could not yet be addressed is whether and how the *Nepenthes* sex chromosomes are involved in divergence with gene flow processes and the evolution of reproductive isolation. This interesting topic, so far mostly studied in animal systems (e.g. *Drosophila*, Presgraves 2008), provides a crosslink between my studies on speciation and sex determination. The combination of genome assemblies from several species with population genomic data would allow to test whether sex chromosomes are protected from introgression and represent hotspots of species divergence in plants, as would be expected based on existing knowledge from animal sex chromosomes.

## References


Adlassnig W, Peroutka M, Lendl T (2011) Traps of carnivorous pitcher plants as a habitat: composition of the fluid, biodiversity and mutualistic activities. *Annals of Botany*, **107**, 181-194.

Athauda SBP, Inoue H, Iwamatsu A, Takahashi K (1998) Acid proteinase from *Nepenthes distillatoria* (Badura). In: *Aspartic Proteinases Advances in Experimental Medicine and Biology*. (ed James MNG), pp. 453-458. Springer US.

Athauda SBP, Matsumoto K, Rajapakshe S et al. (2004) Enzymic and structural characterization of Nepenthesin, a unique member of a novel subfamily of aspartic proteinases. *Biochemical Journal*, **381**, 295-306.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767-780.

Barrett SCH, Hough J (2013) Sexual dimorphism in flowering plants. *Journal of Experimental Botany*, **64**, 67-82.

Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics*, **8**, 382-393.

Brawand D, Wagner CE, Li YI et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375-381.

Charlesworth D (2016) Plant sex chromosomes. *Annual Review of Plant Biology*, **67**, 397-420.

Chin C-S, Peluso P, Sedlazeck FJ et al. (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, **13**, 1050-1054.

Clarke CM (2001) *Nepenthes of Sumatra and Peninsular Malaysia*. Natural History Publications (Borneo), Kota Kinabalu.

Darwin C (1859) *On the origin of species by means of natural selection or the preservation of favored races in the struggle for life*. J. Murray, London.

Delph LF, Knapczyk FN, Taylor DR (2002) Among-population variation and correlations in sexually dimorphic traits of *Silene latifolia*. *Journal of Evolutionary Biology*, **15**, 1011-1020.

Díaz S, Kattge J, Cornelissen JHC et al. (2016) The global spectrum of plant form and function. *Nature*, **529**, 167-171.

Eilenberg H, Pnini-Cohen S, Schuster S, Movtchan A, Zilberstein A (2006) Isolation and characterization of chitinase genes from pitchers of the carnivorous plant *Nepenthes khasiana*. *Journal of Experimental Botany*, **57**, 2775-2784.

Florez-Rueda AM, Paris M, Schmidt A et al. (2016) Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Molecular Biology and Evolution*, **33**, 2935-2946.

Fukushima K, Fang X, Alvarez-Ponce D et al. (2017) Genome of the pitcher plant *Cephalotus* reveals genetic changes associated with carnivory. *Nature* Ecology & Evolution, **1**, 0059.

Hatano N, Hamada T (2008) Proteome analysis of pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteome Research*, **7**, 809-816.

Hatano N, Hamada T (2012) Proteomic analysis of secreted protein induced by a component of prey in pitcher fluid of the carnivorous plant *Nepenthes alata*. *Journal of Proteomics*, **75**, 4844-4852.

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of life reveals clock-like speciation and diversification. *Molecular Biology and Evolution*, **32**, 835-845.

Heubl G, Wistuba A (1997) A cytological study of the genus *Nepenthes* L. (Nepenthaceae). *Sendtnera*, **4**, 169-172.

Knápek O (2012) *Evolution of genome size and DNA base composition in Nepenthes*. Bachelor Thesis. Masaryk University, Brno, Czech Republic.

Lee L, Zhang Y, Ozar B, Sensen CW, Schriemer DC (2016) Carnivorous nutrition in pitcher plants (*Nepenthes* spp.) via an unusual complement of endogenous enzymes. *Journal of Proteome Research*, **15**, 3108-3117.

Mayr E (1942) *Systematics and the origin of species from the view-point of a zoologist*. Columbia University Press, New York.

Moore JC, Pannell JR (2011) Sexual selection in plants. *Current Biology*, **21**, R176-R182.

Nevado B, Atchison GW, Hughes CE, Filatov DA (2016) Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nature Communications*, **7**, 12384.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, Oxford; New York.

Oneal E, Willis JH, Franks RG (2016) Disruption of endosperm development is a major cause of hybrid seed inviability between *Mimulus guttatus* and *Mimulus nudatus*. *New Phytologist*, **210**, 1107-1120.

Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLOS Biology*, **14**, e1002379.

Peng HS, Clarke C (2015) Prey capture patterns in *Nepenthes* species and natural hybrids - are the pitchers of hybrids as effective at trapping prey as those of their parents? *Carnivorous Plant Newsletter*, **44**, 62-79.

Płachno BJ, Adamec L, Lichtscheidl IK et al. (2006) Fluorescence labelling of phosphatase activity in digestive glands of carnivorous plants. *Plant Biology*, **8**, 813-820.

Presgraves DC (2008) Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, **24**, 336-343.

Reich PB, Wright IJ, Cavender-Bares J et al. (2003) The evolution of plant functional variation: traits, spectra, and strategies. *International Journal of Plant Sciences*, **164**, S143-S164.

Rottloff S, Miguel S, Biteau F et al. (2016) Proteome analysis of digestive fluids in *Nepenthes* pitchers. *Annals of Botany*, **117**, 479-495.

Roux J, Privman E, Moretti S et al. (2014) Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, **31**, 1661-1685.

Schraeder CU, Lee L, Rey M et al. (2017) Neprosin, a selective prolyl endoprotease for bottom-up proteomics and histone mapping. *Molecular & Cellular Proteomics*, **16**, 1162-1171.

Wright AE, Darolti I, Bloch NI et al. (2017) Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nature Communications*, **8**, 14251.

Yu Q, Ellen ED, Wade MJ, Delph LF (2011) Genetic differences among populations in sexual dimorphism: evidence for selection on males in a dioecious plant. *Journal of Evolutionary Biology*, **24**, 1120-1127.

# Acknowledgments

This thesis would not have been possible without the help of many people. I express my greatest gratitude to Alex Widmer for giving me the chance to work on *Nepenthes*, his continued scientific interest and support in all matters during this work, and his openness for my own ideas and questions. I thank furthermore my co-examiners Christian Lexer and John Pannell for their efforts and interest in my project.

None of this would have been possible without all my colleagues and office mates in the Plant Ecological Genetics group, who thought me bioinformatics and population genetics, contributed important ideas and discussions, cared for plants, and supported me in countless other ways: Simone Fior, Melanie Schirmann, Niklaus Zemp, Marie Roumet, Martin Fischer, Thomas Städler, Ana Marcela Florez Rueda, Morgane Roth, Margot Paris, Fernanda Baena, Sonja Hassold, James Buckley, Jessica Stapley, Katalin Csilléry, Xuanyu Liu, Alessia Guggisberg, Simon Crameri, Hirzi Luqman, Andre Lohn, Nicolas Defarge, Ursina Messmer, Aksel Pallson, Kevin Vega, Kathrin Rentsch, Adrian Leuchtmann, Artemis Treindl, Monica Garcia, Franziska Berger-Glarner, and Karsten Rohweder. I owe special gratitude to Claudia Michel for introducing me as a novice to the wet lab, her patience and excellent help with all applications. Hirzi Luqman and Christoph Küffer are specially thanked for sharing data from Hirzi's master project on *N. pervillei*. Jean-Claude Walser, Stefan Zoller, Niklaus Zemp, Aria Minder and Silvia Kobel from the Genetic Divesity Center (ETH) helped me with discussions and advise in the wet lab as well as bioinformatics. Peter Hunziker from the Functioncal Genomics Center Zurich was a great help with proteomic analyses.

Essential plant material was contributed by Siegfried and Irmgard Hartmeyer, Michael Zehnder and Urs Zimmermann. They also contributed equally essential, nerdy carnivorous plant discussions, alongside with Ulrike Bauer, Ch'ien C. Lee, Katja Rembold, Bill Hoyer, Shawn Mayes, Charles Clarke, Jeremiah Harris, and many others. This project would never have been initiated without Caroline and Michael Schöner. Peter Enz and the Botanical Garden of the University of Zürich are thanked for contributing leaf material during the early stages of the project. In Singapore, I could not have achieved anything without the support of Lam Weng Ngai. Field work in Brunei Darussalam was certainly among the most enjoyable phases of this project, for which I owe gratitude to Ulmar Grafe and his family, and especially to Aziz, Ieney, and their children in my Telamaba home, as well as Dan Thornham who initially introduced me to tropical field work. Last but not least, my family and friends have supported me with everything during all this time, I could not have done it without you.

# Curriculum Vitae

## Mathias Scharmann

born 09. March 1986 in Beckum, Germany

**Education**

| | |
|---|---|
| 29.06.2017 | passed doctoral examination: "Evolutionary genomics of carnivorous *Nepenthes* pitcher plants" |
| 12.2012 - 06.2017 | doctoral student, ETH Zürich, supervisor Alex Widmer |
| 10.01.2012 | Diplom-Biologe Univ., Universität Würzburg mark 1.0 (best) |
| 01.2011 - 11.2011 | diploma thesis, supervisor W. Federle at University of Cambridge: "Biomechanical and ecological adaptations of the pitcher plant ant *Camponotus schmitzi*" |
| 10.2006 - 11.2010 | studies in Biology, University of Würzburg, Germany |

**Peer-reviewed Publications**

(4) Lim, Y.S., Schöner, C.R., Schöner, M.G., Kerth, G., Thornham, D.G., Scharmann, M. & Grafe, T.U. (2015) How a pitcher plant facilitates roosting of mutualistic woolly bats. *Evolutionary Ecology Research* 16, 581-591.

(3) Bauer, U., Scharmann, M., Skepper, J. & Federle, W. (2013) 'Insect aquaplaning' on a superhydrophilic hairy surface: how *Heliamphora nutans* Benth. pitcher plants capture prey. *Proceedings of the Royal Society B* 280 (1753): 20122569.

(2) Scharmann, M., Thornham, D., Grafe, T.U. & Federle, W. (2013) A novel type of tripartite ant-plant interaction: Ant partners of carnivorous pitcher plants prevent nutrient export by dipteran pitcher infauna. *PLOS ONE* 8, e63556.

(1) Scharmann, M. & Grafe, T.U. (2013). Reinstatement of *Nepenthes hemsleyana* (Nepenthaceae), an endemic pitcher plant from Borneo, with a discussion of associated *Nepenthes* taxa. *Blumea* 58, 8-12.

**Other Publications**

Scharmann, M. (2014). Special among a special flora - the carnivore *Nepenthes pervillei*. *Kapisen* 17, 18-19.

Grafe, T.U., Goutte, S., Ahmadsah, H.H., Konopik, O., Scharman, M. & Bauer, U. (2010). Updated check list of the amphibians of the Ulu Temburong National Park and Brunei Darussalam. *Scientia Bruneiana* 2010 Issue, 53-59.

Scharmann, M. (2010). Der Pulong Tau Nationalpark - Reise ins Herz von Borneo. *Das Taublatt* 67 & 68, 52-77 & 40-65.

typeset in LaTeX inspired by the typographical look-and-feel `classicthesis`