New Phytologist

Research

Rapid radiations of both kiwifruit hybrid lineages and their parents shed light on a two-layer mode of species diversification

Yifei Liu[1]*, Dawei Li[2]*, Qiong Zhang[2]*, Chi Song[3]*, Caihong Zhong[2]*, Xudong Zhang[3], Ying Wang[3], Xiaohong Yao[2], Zupeng Wang[1], Shaohua Zeng[1], Ying Wang[1], Yangtao Guo[1], Shuaibin Wang[1], Xinwei Li[2], Li Li[2], Chunyan Liu[2], Honour C. McCann[1,4], Weiming He[1], Yan Niu[3], Min Chen[3], Liuwen Du[2], Junjie Gong[2], Paul M. Datson[5], Elena Hilario[5] and Hongwen Huang[1,2]

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization and Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, The Chinese Academy of Sciences, Guangzhou, Guangdong 510650, China; [2]Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, The Chinese Academy of Sciences, Wuhan, Hubei 430074, China; [3]Wuhan Benagen Tech Solutions Company Limited, Wuhan, Hubei 430070, China; [4]New Zealand Institute for Advanced Study, Massey University, Auckland 0745, New Zealand; [5]The New Zealand Institute for Plant and Food Research Limited, Mt Albert Research Centre, Auckland 1142, New Zealand

Author for correspondence:
Hongwen Huang
Tel: +86 20 37252778
Email: huanghw@scbg.ac.cn

Received: *30 November 2016*
Accepted: *4 April 2017*

New Phytologist (2017)
doi: 10.1111/nph.14607

Key words: *Actinidia*, backbone phylogeny, hybrid speciation, phylogenomics, reticulate evolution.

Summary

- Reticulate speciation caused by interspecific hybridization is now recognized as an important mechanism in the creation of biological diversity. However, depicting the patterns of phylogenetic networks for lineages that have undergone interspecific gene flow is challenging.
- Here we sequenced 25 taxa representing natural diversity in the genus *Actinidia* with an average mapping depth of 26× on the reference genome to reconstruct their reticulate history.
- We found evidence, including significant gene tree discordance, cytonuclear conflicts, and changes in genome-wide heterozygosity across taxa, collectively supporting extensive reticulation in the genus. Furthermore, at least two separate parental species pairs were involved in the repeated origin of the hybrid lineages, in some of which a further phase of syngameon was triggered. On the basis of the elucidated hybridization relationships, we obtained a highly resolved backbone phylogeny consisting of taxa exhibiting no evidence of hybrid origin. The backbone taxa have distinct demographic histories and are the product of recent rounds of rapid radiations via sorting of ancestral variation under variable climatic and ecological conditions.
- Our results suggest a mode for consecutive plant diversification through two layers of radiations, consisting of the rapid evolution of backbone lineages and the formation of hybrid swarms derived from these lineages.

Introduction

Understanding the patterns and mechanisms underlying biological diversification remains central to answering questions about life on Earth. The classical phylogenetic foundation from which diversity grows in a bifurcating tree-like pattern has frequently been questioned, particularly in plants, where interspecific hybridization is prevalent (Rieseberg & Willis, 2007; Soltis & Soltis, 2009). Increasing evidence, from species divergence accompanied by gene flow to hybrid speciation, is driving the emergence of network models of diversification (Nakhleh, 2013), reflecting a dynamic process of organism evolution with genetic exchanges (Jónsson *et al.*, 2014; Lamichhaney *et al.*, 2015; Leducq *et al.*, 2016; Mallet *et al.*, 2016; Pease *et al.*, 2016). A growing body of genome-scale analyses further provides convincing examples showing reticulations across the 'Tree of Life', including analyses tracking clues of ancient introgression (Green *et al.*, 2010) and clarifying the evolutionary order of taxa with extensive interspecific gene flow (Heliconius Genome Consortium, 2012; Fontaine *et al.*, 2015), lending increasing support to the validity of the 'Web of Life' metaphor (Arnold, 2015). However, current data demonstrating the validity of reticulate diversification may represent just the tip of the iceberg given the potential for such diversification to be discovered in all domains of life.

Despite widespread interest in evolutionary networks, there remains little consensus among evolutionary biologists regarding the extent and pathways of gene flow occurring during speciation and diversification (Burke & Arnold, 2001; Mallet, 2007; Mavárez & Linares, 2008). In plants, allopolyploidy is a well-established speciation mode, while hybrid speciation without changes in ploidy levels (homoploid hybrid speciation) is less

*These authors contributed equally to this work.

clearly established (Mallet, 2007; Soltis & Soltis, 2009). Given that reduced contact and competition with parental lineages coincide with the occurrence of new ecological opportunities, a hybrid swarm theory predicts that gene flow can act as a promoter, catalyzing hybrid speciation and rapid radiations (Seehausen, 2004). The volume of diversity acquired in a rapid radiation can be further elevated when occasional or locally restricted hybridization occurs between diverging units within the radiation; this phase is considered to be a syngameon, in which the most inclusive unit of interbreeding is represented by a hybridizing species group (Grant, 1981). In this view, the increased genetic variation produced through hybridization and the unexploited resources or ecological niches available for hybrid genotype colonization are both critical (Yakimowski & Rieseberg, 2014). Environmental change, therefore, has been thought of as a particularly efficient driver of hybrid speciation in that it promotes both the secondary contact of separated evolutionary lineages and the emergence of new niches (Nolte & Tautz, 2010). However, these hybrid events are technically much more difficult to trace. Moreover, the genetic sources of the populations that take advantage of new ecological opportunities are possibly diverse (Pease *et al.*, 2016), with radiations of both the parental lineages and their hybrid progenies likely to be taking place simultaneously, giving rise to rapid consecutive evolution of life diversity (Gavrilets & Losos, 2009).

Kiwifruit (*Actinidia* Lindl.), a genus of economically valuable fruiting vine species, is known for its high morphological and ecological variation. More than 50 species have been described (Li *et al.*, 2007), which fall into two distinct groups: those with smooth-skinned fruit (SSF) and those with hairy and/or spotted (lenticillate) fruit (HSF). The genus has a remarkably wide geographic distribution in eastern Asia, extending from the Equator (tropics) to cold temperate regions, with the vast majority of taxa occurring in the mountains and hills of south central and southwestern China (Huang, 2014). The SSF group tends to occur in cold and high-altitude areas, mainly in central-western and northern China, while the HSF group is found in relatively warm and moist environments in southern and eastern China. Molecular studies using multi-locus methods have provided supporting information on the monophyly of the genus as a whole but revealed contradictory relationships between *Actinidia* taxa, particularly among members of the HSF group (Li *et al.*, 2002; Chat *et al.*, 2004). Intrageneric sections within *Actinidia* have been defined but are of questionable value as there is overlap of traits between taxa (Huang, 2014). Natural hybridization is thought to be the largest contributor to phylogenetic incongruities (Li *et al.*, 2002; Chat *et al.*, 2004; Huang, 2014). Indeed, many *Actinidia* taxa can be crossed to produce hybrids under experimental conditions, suggesting that there is incomplete reproductive isolation. Moreover, wild *Actinidia* populations frequently consist of several sympatric taxa (Huang, 2014), providing opportunities for interspecific gene flow. However, because of the paucity of genome sequences for *Actinidia* species, conclusive evidence for the presence and extent of hybridization among *Actinidia* taxa is lacking. Questions with respect to the evolutionary trigger and pattern responsible for the high species and ecological diversity of this genus remain to be answered.

Phylogenetic reconstruction of lineages that have undergone reticulation is notoriously challenging because reticulation events result in genomic regions with local genealogies that are incongruent with the speciation pattern (Linder & Rieseberg, 2004). This difficulty can be further compounded by the sorting of ancestral polymorphisms as a consequence of incomplete lineage sorting (ILS) (Whitfield & Lockhart, 2007), which is especially common in groups radiating both in recent history and at longer timescales (Pease *et al.*, 2016). In such cases, scanning entire genomes of closely related organisms is a powerful way to detect genome-wide signals of heterogeneity and to unravel the reticulate history of organism evolution (Posada, 2016).

Here, we generated the first whole-genome data set consisting of 40 *Actinidia* samples, representing 25 taxa in both the SSF and HSF groups, as well as one sample produced in a breeding program (referred to as MTH) (Supporting Information Notes S1). We used this genomic information to detect the presence of hybridization among *Actinidia* taxa at an unprecedented level of detail, with our analyses including determination of the extent and direction of hybridization involved in a particular reticulate event, and investigation of the mosaic genome structure of a hybrid lineage inherited from its putative donor parents. We also refined a rapidly radiated evolutionary backbone phylogeny underlying extensive reticulation and hybrid speciation in *Actinidia*, and argue that the reticulate pattern identified in this genus may represent a potential universal mode of triggering rapid plant diversification.

Materials and Methods

Plant material, sequencing and variant discovery

We obtained plant material from 34 samples of 23 *Actinidia* Lindl. taxa from the National *Actinidia* Germplasm Repository (NAGR; Wuhan, China) which had been originally collected from the wild, with material from an additional four samples of *Actinidia arguta* (Siebold and Zuccarini) Planchon ex Miquel and a sample of the outgroup *Clematoclethra* (Franchet) Maxim. provided by the Xi'an Botanic Garden, Xi'an, China. We also included the Illumina (Illumina Inc., San Diego, CA, USA) sequences of two samples of *Actinidia hypoleuca* Nakai and *Actinidia melanandra* Franchet, provided by Plant & Food Research, Auckland, New Zealand (Figs S1, S2; Table S1). Total genomic DNA was isolated using the cetyltriethylammnonium bromide (CTAB) method (Doyle & Doyle, 1987) with minor modifications. At least 5 μg of genomic DNA from each sample was used to construct a sequencing library following the manufacturer's instructions (Illumina Inc.). Paired-end sequencing libraries with insert sizes of *c.* 180 bp and/or 500 bp were sequenced using the Illumina HiSeq 2000 platform. Raw sequencing data were filtered using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/).

We used Stampy v.1.0.21 to map the paired-end resequencing reads for all taxa to the kiwifruit reference genome (Huang *et al.*,

2013). The aligned reads were filtered to remove PCR duplicates and sorted according to mapping coordinates with the PICARD package v.1.93 (http://picard.sourceforge.net/). We further realigned the circum-indel regions using the Genome Analysis Toolkit (GATK) v.3.0-0-g6bad1c6 (McKenna *et al.*, 2010). The resulting alignment files were subject to genotyping using the GATK UnifiedGenotyper at each reference locus. We applied a series of stringent filtering processes to the potential variants to produce the final set of single nucleotide polymorphisms (SNPs) (Notes S1; Figs S3, S4). To identify organellar genomic SNPs, BWA v.0.7.4-r385 (Li & Durbin, 2009) was used to map reads on the available chloroplast (cp) reference genome (Yao *et al.*, 2015) and a partially assembled mitochondrial (mt) reference genome (Notes S1; Fig. S5) in *Actinidia*. We used a similar filtering process in nuclear SNP calling to obtain the final SNP sets on cp and mt genomes separately.

For polyploid samples, SNP identification in nuclear genomes is challenging as a consequence of the presence of both homeologous SNPs (polymorphic positions occurring across subgenomes within and among individuals) and true allelic SNPs (polymorphic positions occurring within a single subgenome among individuals) (Clevenger *et al.*, 2015). Without extra high coverage depth of mapping reads on a good reference genome of sufficient quality, it is impossible to unambiguously distinguish between homeologous SNPs and true allelic SNPs. We thus mapped reads from all subgenomes of a polyploid to the reference and considered the two types of SNP together as 'mixed SNPs' in subsequent analyses. This approach would be reliable at least for detecting hybridization between *Actinidia* species, where a polyploid hybrid would show a particularly high level of genome-wide heterozygosity (Fig. S6). Moreover, as the majority of SNPs found in kiwifruit polyploids were biallelic (Fig. S7), we filtered out tri- and tetra-allelic SNPs and only used high-quality SNPs (mapping quality >30; depth >10) to further reduce possible bias in phylogenomic analysis (Notes S1).

Phylogeny inference and gene tree analysis

Maximum likelihood (ML)-based phylogenetic trees of nuclear, cp and mt genomes were constructed using RAXML v.8.1.15 (Stamatakis *et al.*, 2005). We used a single substitution model of GTR+GAMMA and 500 bootstrap replicates for all the different analyses in different genomic types. The correlated rate model in MCMCTREE in the PAML package v.4.8a (Yang, 2007) was used for dating the backbone phylogeny based on SNP sites with an approximate likelihood calculation (Notes S1).

We performed gene tree analysis using the program BUCKY v.1.4.0 without making any assumptions as to the source of the discordance. We filtered 100-kb nonoverlapping windows across the genomic alignments as 'genes' for tree constructions. We used MRBAYES v.3.2.5 to obtain a posterior distribution of gene trees for these genomic windows. Two Markov chain Monte Carlo (MCMC) runs were performed with one cold chain and three heated chains for chain length (and burn-ins) of 1000 000 (10 000) and sampled every 100 generations. We then used BUCKY to construct both a primary concordance tree from the clades with the largest concordance factors (CFs), to capture the main vertical phylogenetic signals, and an estimated population tree inferred from quartet CFs. For each set of four taxa, there are three possible quartets. The quartet with the greatest CF is retained, and a population tree is built from this set of quartets using the quartet-joining algorithm (Xin *et al.*, 2007). This quartet-based method is guaranteed to recover the true species tree if all discordance is attributable to the coalescent process along the tree (Larget *et al.*, 2010). We used a CF cutoff value of ≥ 0.05 to investigate alternative bipartitions from the primary tree for potential evidence of hybridization or ILS (Notes S1).

Identifying signal of hybridization

We followed the methodology described in Wu *et al.* (2014) to calculate the genome-wide heterozygosity of each *Actinidia* sample sequenced. For polyploids, we used the mixed types of SNPs to calculate the genome-wide heterozygosity (Notes S1).

We used an approximately unbiased (AU) test (Shimodaira, 2002) to investigate support for the three possible topologies of an unrooted four-taxon tree (e.g. S_1, S_2, S_3 and O; $S_{1–3}$ are ingroup kiwifruit taxa subjected to a comparison, and O is an outgroup sample) in 10-kb nonoverlapping windows along kiwifruit genomes. A tree topology with an AU *P*-value > 0.95 means strong support of that topology from the alignment. Partitions of alignment with an observed likelihood difference of 0 were excluded accordingly, and significant asymmetry in support for the three topologies was determined by calculating 95% confidence intervals using 1000 nonparametric bootstrapping replicates in R v.3.2.2 (https://www.r-project.org/).

We also calculated Patterson's *D* statistic (Green *et al.*, 2010; Durand *et al.*, 2011) for the same four-taxon alignments using modified scripts originally used for similar analysis in butterflies (Martin *et al.*, 2013). For the ordered alignment (S_1, S_2, S_3 and O), we called the two allelic configurations of interest 'ABBA' or 'BABA'. The ABBA site refers to a pattern where S_1 has the outgroup allele and S_2 and S_3 share the derived copy. The BABA site corresponds to patterns where S_1 and S_3 share the derived allele and S_2 has the outgroup allele. Under the null hypothesis of ILS, the number of ABBA and BABA sites is expected to be equal ($D = 0$). Alternatively, significant deviation of *D* from 0 suggests other events, in particular S_3 exchanging genes with S_1 or S_2 (hybridization on which we focused) (Durand *et al.*, 2011). Conservatively, only sites at which the four-taxon genomes were homozygous for the same allele were considered to ensure confident assignment of the ancestral and derived states across species.

Genomic regions sharing identity by descent (IBD) between diploid samples were identified using BEAGLE v.4.1 (https://faculty.washington.edu/browning/beagle/beagle.html), and degrees of relatedness between them were investigated using the KING program (Manichaikul *et al.*, 2010).

Investigating parental contributions

We developed a strategy by using both the AU test and Patterson's *D* statistic to investigate parental contributions to the

mosaic genome of an identified hybrid. For a four-taxon alignment (P1, P2, H and O), P1 and P2 were both donor species in relation to a derived hybrid (H) with an outgroup (O). The percentage of AU tests supporting both tree topologies (((P1, H), P2), O) and (((P2, H), P1), O) would reflect the respective genomic contribution from each parent (Fig. S8). If the genomic contributions from the two parents were nearly equivalent, we expected that the AU test percentages supporting the two gene tree topologies would be similar (both *c.* 50%). Moreover, for the same alignment, Patterson's *D* statistic would show the status of both 'ABBA' (P2 and H) and 'BABA' (P1 and H) sites. If a non-significant *D* statistic value was found ($|Z|$ scores < 3), it would indicate a balance of 'ABBA' and 'BABA' sites, reflecting a similar level of genomic contribution from each donor parent. Alternatively, asymmetric parental contributions would be revealed by a significant *D* statistic.

Genetic divergence, core genome and assembly of unmapped reads

Genetic divergence between diploid genomes was characterized by a D_d value (*D* value in Wu *et al.* (2014); we use D_d here to distinguish it from Patterson's *D* statistic) in 100-kb genomic windows sliding 50 kb at a time. We considered a core genome region if it had sequence coverages of more than one read in each of the backbone taxa based on the BAM files. The core genes shared were determined by examining the coding sequences of intact gene structure in the core genome regions. We also used SAMtools v.0.1.18 (Li *et al.*, 2009) to obtain the unmapped reads from BAM files of each backbone taxa, and used SOAPdenovo2 v.2.04 (Luo *et al.*, 2012) to assemble reads. For scaffolds longer than 2 kb, we used Maker v.2.31.8 (http://www.yandell-lab.org/software/maker.html/) for gene predictions. All gene ontology (GO) annotations were performed using Blast2GO v.2.5.0 (Conesa *et al.*, 2005) based on the $Blast_P$ results ($Blast_P$, *e*-value of 1e-5) of the nr database (20160801). GO enrichment analysis for the annotated genes in the putative sweep regions was performed using Goseq v.1.16.2 (Young *et al.*, 2010). GO terms with Benjamini–Hochberg corrected *P*-values < 0.05 were considered significantly enriched (Notes S1).

Demography and ecological niche modeling

We used the multiple sequentially Markovian coalescent (MSMC) approach (Schiffels & Durbin, 2014) to draw inferences about changes in the past effective population size (N_e). The MSMC is an extension of the pairwise sequentially Markovian coalescent (PSMC) method (Li & Durbin, 2011). When one individual is used, the MSMC and PSMC analyses produce similar results, but the former is potentially more accurate (Schiffels & Durbin, 2014). Results of demographic modeling were scaled using an assumed neutral mutation rate of 1×10^{-8} per bp per generation in higher plants (Gaut, 1998) and a generation time of 7 yr (Huang, 2014).

We modeled ecological niche distribution over time using MaxEnt v.3.3.3k (Phillips *et al.*, 2006). Nineteen bioclimatic variables were retrieved from the WorldClim database (http://www.worldclim.org/) to perform modeling, and pairwise Pearson's correlation was used to reject highly correlated variables (≥ 0.8). The jack-knife test and the area under the curve (AUC) with training and test data were also calculated in MaxEnt to evaluate the accuracy of the models. We predicted changes of the distribution of these backbone species between three time periods: the present day, the last glacial maximum (LGM) (*c.* 0.022 Ma), and the Mid-Holocene (MH) (*c.* 0.006 Ma). We used both LGM and MH data simulated by the Model for Interdisciplinary Research on Climate (MIROC) (Hasumi & Emori, 2004).

Results

Sequencing and a genome-based phylogeny

We sequenced *c.* 0.8 trillion DNA reads for all *Actinidia* samples. We aligned reads to the genome assembly of an *Actinidia chinensis* individual (Huang *et al.*, 2013), yielding an average depth of *c.* 26× per species (Table S2). The average genome coverage was *c.* 77–96% of the reference for all in-group samples. We applied stringent filters (Notes S1) to identify a final set of 7.89 million SNPs. Among these, 643 620 nonsynonymous SNPs in 31 718 genes were identified, including 513 651 missense and 19 385 nonsense variants (Table S3). Unless otherwise noted, we used a subset of 30 samples without intraspecific phylogenetic redundancies (Notes S1) to perform all subsequent analyses.

We produced an ML phylogeny using all SNPs as a total evidence approximation of the likely species relationships. The resulting total evidence nuclear (TEN) tree (hereafter referred to as the 'concatenated tree') was resolved for the majority of clades (Figs 1a, S9). The SSF and HSF groups are monophyletic with full (100%) bootstrap support. The TEN concatenated tree revealed a subdivision within the SSF group, recovering two sister clades: the *Actinidia arguta* complex and the *Actinidia polygama* clade. The taxa and clades within the HSF group are paraphyletic, with a basal clade of the *A. chinensis* complex. The tree topography was generally congruent with previous reports based on nuclear internal transcribed spacer (ITS) fragments (Li *et al.*, 2002) and organellar genes (Chat *et al.*, 2004). However, more important deviations than were previously thought to exist were observed, such as the monophyly of the SSF group and the relatively basal positions of both the *A. chinensis* clade and *Actinidia rufa* within the HSF group.

Incongruence between gene trees and species phylogeny

To investigate any conflicting signals in relation to gene trees and species phylogeny, we constructed trees separately from each 100-kb nonoverlapping genomic window (referred to here as 'gene trees' regardless of their protein-coding content; $n = 3691$ windows) (Figs 1b, S10; Notes S1). Using Bayesian concordance analyses (BCAs) (Larget *et al.*, 2010), we obtained a primary concordance tree (referred to hereafter as a 'species tree') representing

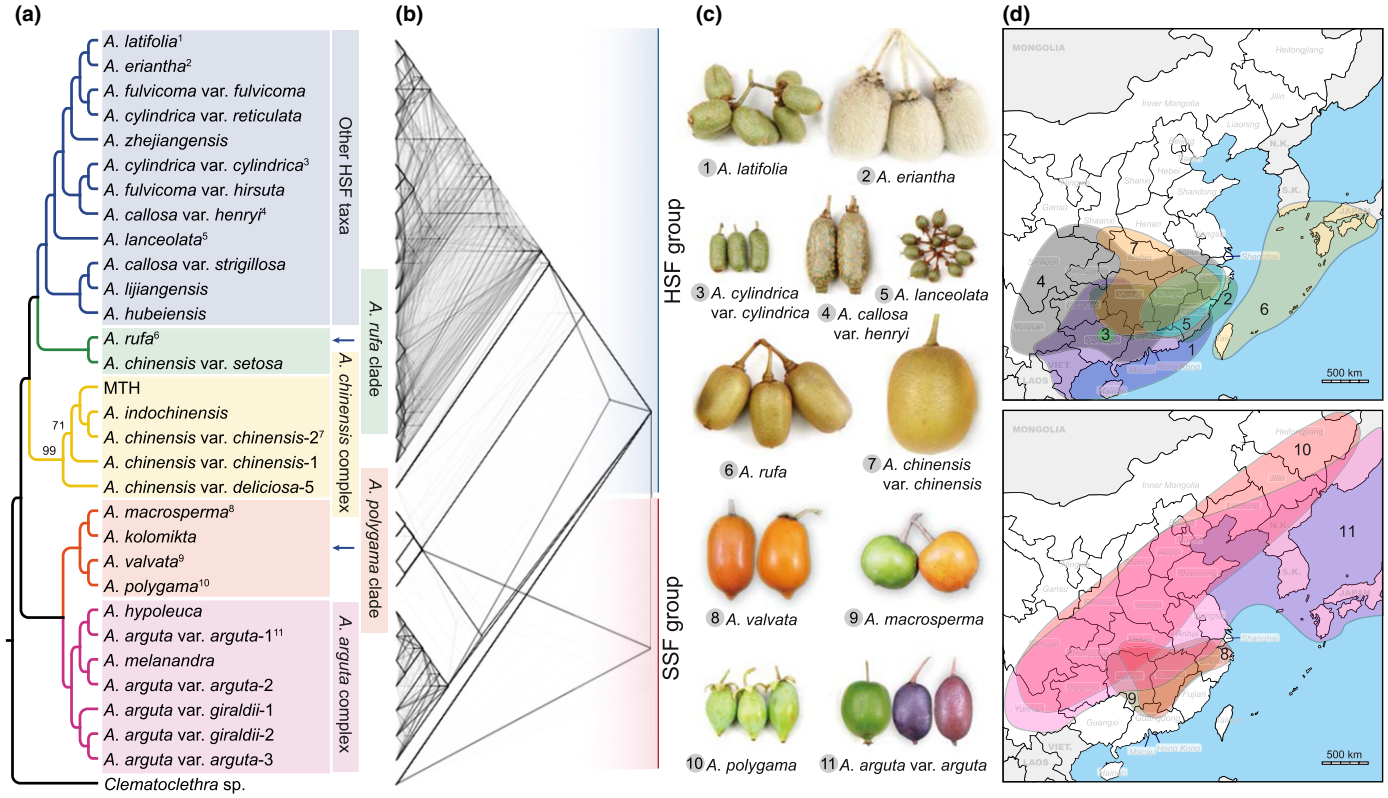

Fig. 1 Genome-scale phylogenetic relationships of sequenced *Actinidia* taxa. (a) The total evidence nuclear (TEN) concatenated tree constructed using genome-wide single nucleotide polymorphisms (SNPs). All nodes have 100% bootstrap support except those two nodes noted on the tree. Main species complexes and clades are highlighted. (b) The overlapping gene trees from 3691 genomic windows. (c) Fruits of representative *Actinidia* taxa, and (d) their possible natural range of distribution. Their phylogenetic positions are indicated by superscript numbers in the TEN concatenated tree.

the sample- or genome-wide concordance (annotated with a CF, an estimate of the proportion of the sampled genes or the genome for which the clade is true) of all gene trees constructed, and also a population coalescent tree (Fig. S11). Surprisingly, we found incongruence between the BCA species tree and the TEN concatenated tree, as well as between the population tree and the concatenated tree (Fig. S11), reflecting contradictory results in terms of clades differing between gene trees, including between individual gene trees and the species phylogeny (Fig. 1).

We investigated major discordance between individual gene trees and the species tree. Employing a CF cutoff of 5%, we identified discordances on both deep and terminal branches (Fig. 2a). For deep branches, the three main clades of the whole tree (the HSF group, the *A. polygama* clade, and the *A. arguta* complex clade) revealed similarly high CFs (0.203–0.401) to each other (Fig. S12). The symmetry in CFs was consistent with conditions that could lead to ILS, a population-level process resulting in incongruence between gene trees and the species tree (Maddison, 1997). We also identified three anticipated ILS patterns on relatively terminal branches constructed from samples within a species complex, including the *A. arguta* and the *A. chinensis* complexes, or from a group consisting of very closely related HSF taxa (Fig. S12).

We further found evidence of extensive asymmetric discordance resulting in certain taxon or clade pairs frequently grouping together (Fig. 2). Two members of the SSF clade (*A. polygama* and *Actinidia valvata*) exhibited the highest CF of 0.432 (Fig. 2a), reflecting a very high proportion (43.2%) of sampled genes supporting the only cluster of *A. polygama* and *A. valvata*. Similarly, many taxa were linked to two other distinct lineages simultaneously and asymmetrically (Fig. 2b–f). For example, *Actinidia fulvicoma* var. *fulvicoma* showed genetic similarity to *A. cylindrica* var. *reticulata*, and these taxa were further linked to *Actinidia eriantha* (and its most recent common ancestor (MRCA)), and *A. cylindrica* var. *cylindrica* (and its MRCA), respectively (Fig. 2b). Similar patterns were found for the three interrelated taxa *Actinidia lijiangensis*, *Actinidia callosa* var. *strigillosa*, and *Actinidia hubeiensis* (Fig. 2c) and also for *Actinidia zhejiangensis* (Fig. 2d), and *A. chinensis* var. *deliciosa*-5 (Fig. 2e). The control sample (MTH) clustered with *A. chinensis* but showed a relationship with the *A. eriantha* lineage (Fig. 2f), consistent with the recorded introgressive parent of *A. eriantha* (Notes S1). These relationships suggested that reticulation, as opposed to ILS alone, contributed to kiwifruit evolution.

Lines of evidence rigorously support extensive hybridization

Actinidia displays paternal and maternal transmission of the cp and mt genomes, respectively (Testolin & Cipriani, 1997; Chat *et al.*, 1999), allowing the reconstruction of organellar phylogenies which are valuable for detecting reticulation. Besides the

New Phytologist

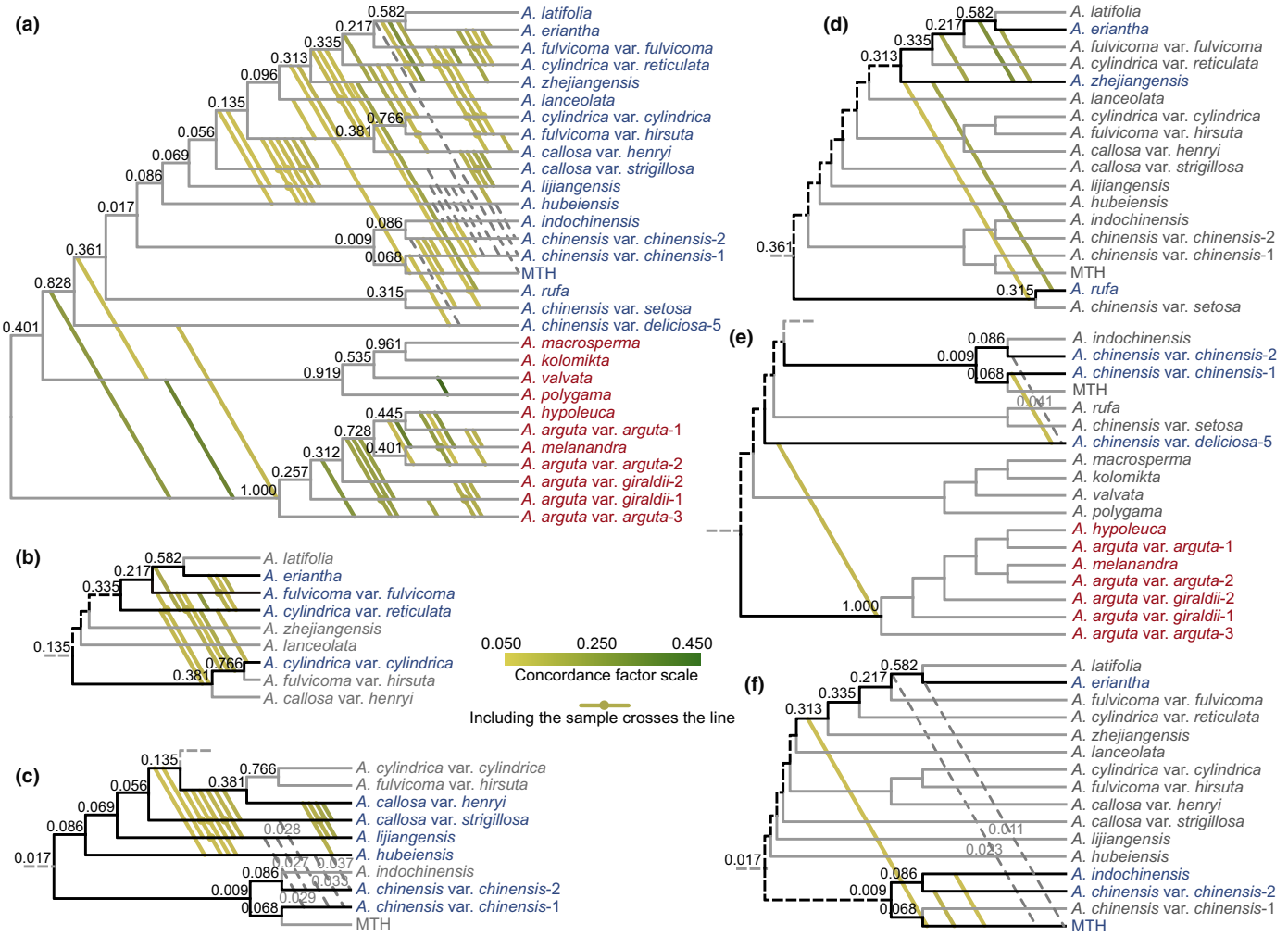

Fig. 2 Discordances between gene trees in relation to reticulation. (a) All significant discordances (concordance factor (CF) ≥ 5%) identified on the Bayesian concordance analysis (BCA) species tree. (b, c) Complex asymmetric discordances show that hybrid formation repeatedly occurred between members of one pair of donor species. (d–f) Alternative cases of hybridization. Discordances with CFs ≥ 5% are indicated with solid lines with color gradients, while those having CFs < 5% but with gene flow detected are indicated with dashed lines. The smooth-skinned fruit (SSF) and hairy and/or spotted (lenticillate) fruit (HSF) samples are indicated in red and blue, respectively.

available cp reference genome in *Actinidia* (Yao *et al.*, 2015), we assembled an mt reference genome from our whole-genome sequencing data, resulting in a partial assembly with a total length of 332 117 bp (Notes S1). We mapped all resequencing reads to the cp and mt reference genomes, generating respective read coverage depths of at least 105× and 201× for each genome (Tables S4, S5). Using 3269 genome-wide cp SNPs and 7291 mt SNPs, we constructed both cp and mt genomic trees. To determine the presence and impact of hybridization, we explored cytonuclear conflicts as evidence of interspecific gene flow. We found a low concordance of branching patterns between nuclear, cp and mt genomic trees (Fig. S13). With precise identification of conflicting signals, we found clear evidence of cytonuclear conflict: ten samples (including MTH) producing six different combinations of cp, mt and nuclear types (Fig. 3a; a type was defined by the most closely related species of a sample investigated on the respective cp, mt and nuclear genomic trees).

We also investigated nuclear genome-wide heterozygosity as evidence of hybridization. Notably, 11 samples (*Actinidia kolomikta* and the 10 described above displaying cytonuclear conflicts) showed substantially higher genome-wide heterozygosity than other samples at the same ploidy level (Figs 3b, S6, S14; Table S6). The control, MTH, also exhibited increased heterozygosity, in which the first two peaks corresponded to the heterozygosity of its hybrid parents, and the third peak (on the extended right tail; Fig. 3b) reflected the heterozygosity of the F_1 offspring left before backcrossing. The fact that the taxa with elevated heterozygosity also exhibited asymmetric discordance in relation to specific species or clades on the BCA species trees (Fig. 2) suggests that they originated via hybridization.

The genome-wide data allowed us to perform rigorous tests for gene flow in the face of ILS. The AU test (Shimodaira, 2002) and Patterson's *D* statistic (Green *et al.*, 2010) analyses were

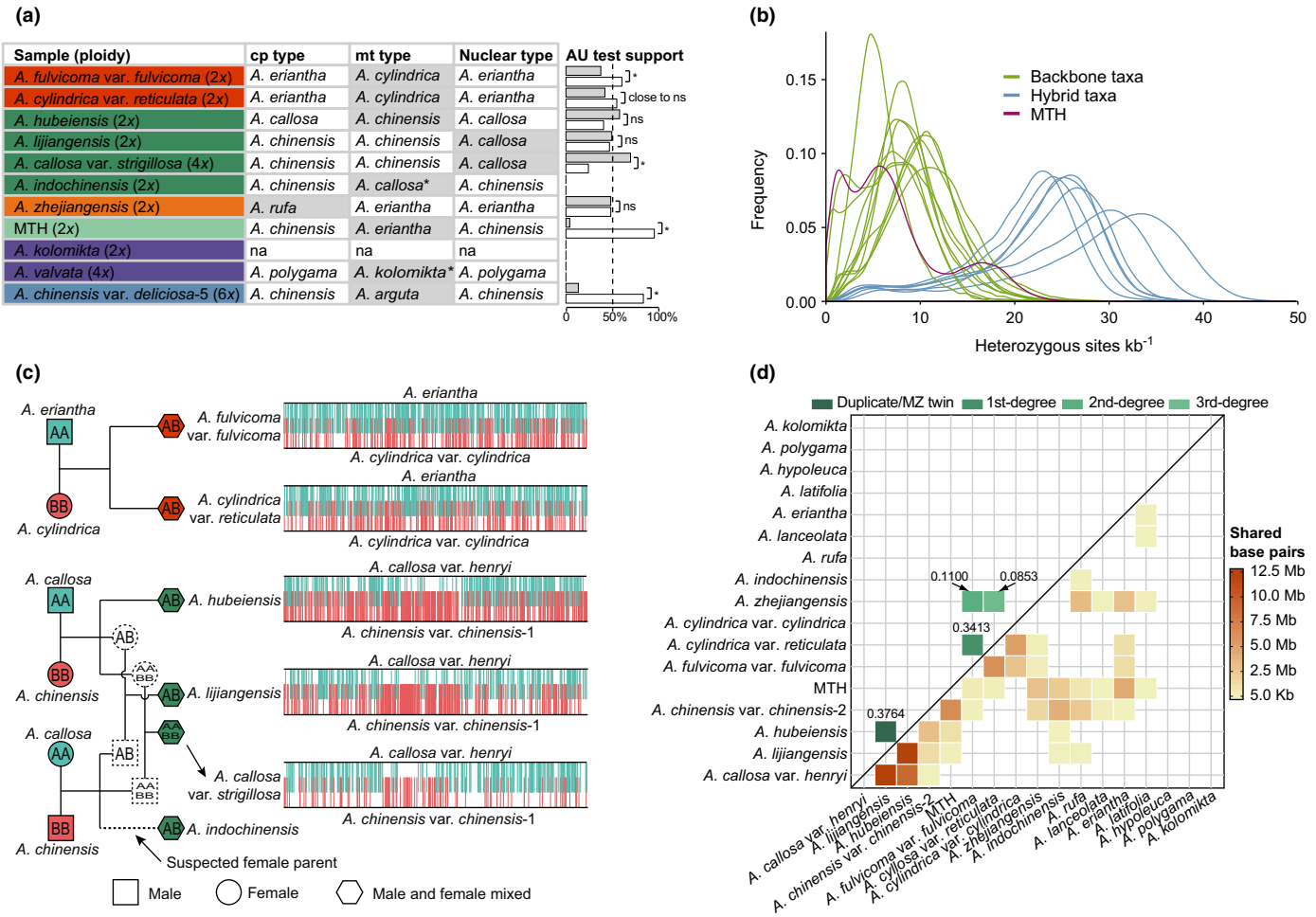

Fig. 3 Evidence supporting extensive and repeated hybridization and hybrid speciation. (a) *Actinidia* hybrids inferred from cytonuclear conflicts and/or increased genome-wide heterozygosity. *, parental type suspected; NA (not available), parental type undetermined (see Supporting Information Notes S1 for further discussion). Hybrid samples with the same color have the same putative parents. The AU test and Patterson's *D* statistic analyses showing parental genomic contributions are also presented (Table S8). Significant deviation in the *D* statistic (Z score > 3) is denoted by an asterisk, while others are nonsignificant (ns). (b) The relative distribution of nucleotide heterozygosity among diploid samples. (c) Cases of repeatedly derived hybrids with distinct mosaic genomes (chromosome 1; see Fig. S16 for all chromosomes). Open circles or squares with dashed lines indicate putative intermediate hybrids within a syngameon scenario. (d) Estimated genetic relatedness (upper triangle) and genomic regions that share identity by descent (IBD) (lower triangle) of diploid *Actinidia* taxa. cp, chloroplast; mt, mitochondria; MTH, a sample produced in a breeding program.

conducted for all extant taxa showing significant gene tree discordance (CF > 0.05). Both analyses employ a four-taxon tree topology, using which the AU test investigates significant asymmetry in support of a clustering pattern of the three in-group taxa, while the *D* statistic determines a statistically significant imbalance in the number of sampled discordant biallelic site patterns 'ABBA' and 'BABA', providing evidence that hybridization has occurred. The results of 11 of 14 four-taxon tests (Table S7) were attributed to hybridization as they exhibited both asymmetry in the AU test and significant deviation of *D* values from 0 (|Z|-score > 3). All of these cases were in relation to the asymmetric discordances which were showed in the Fig. 2(b–f). We then included four additional tests with asymmetric relationships despite low CFs (< 0.05) (Table S7). We found that all of them were associated with gene flow between the taxa in question. By contrast, we confirmed anticipated ILS between the *A. arguta* complex and the *A. polygama* clade ($D = 0.01 \pm 59\%$; Z score = 2.46; $P = 0.07$) (Fig. S12) and found that ILS rather than reticulation was exhibited between *A. callosa* var. *henryi* and *A. fulvicoma* var. *hirsuta*, and between *A. chinensis* var. *setosa* and *A. chinensis* var. *chinensis*-2 (Fig. S15; Table S7).

Syngameon and repeated hybrid speciation

Interestingly, we found events in the repeated origin of hybrid lineages in relation to at least two separate parental species pairs of *A. eriantha* and *A. cylindrica*, and *A. chinensis* and *A. callosa* (Fig. 3), resulting in multiple hybrid taxa being successfully established. We further found that two taxa, *A. lijiangensis* and *A. callosa* var. *strigillosa*, had both cp and mt types of a single parent, *A. chinensis*, but the nuclear type of the other parent, *A. callosa* (Fig. 3), suggesting more complex evolutionary scenarios involving multiple hybridization events. We suspected that scenarios existed in which there were potentially two steps of

hybridization (and polyploidy) with putative intermediate hybrid lineages involved (Fig. 3c).

We developed a method on the basis of the AU test and Patterson's *D* statistic analyses to investigate the mosaic genome structure of a hybrid inherited from its putative parents. Given a four-taxon alignment (P1, P2, H and O) in which P1 and P2 were both donor species in relation to a derived hybrid (H) with an outgroup (O), the percentage values from the AU test supporting each of the gene tree topologies (((P1, H), P2), O) and (((P2, H), P1), O) would reflect the relative genomic contribution from each parent (Fig. S8), and further, the *D* statistic showing the genome-wide status of 'ABBA' and 'BABA' sites was consistent with the genomic similarities between P2 and H, and P1 and H, respectively. For eight hybrid taxa with donor parents precisely identified, we found that four tests had nearly symmetric AU test results supporting both gene trees (between 41% and 55%), while the other four did not (Figs 3a, S16; Table S8). We validated three of the four tests having symmetric tree supporting data and found that they did not show significant deviation in the *D* statistic (Z score < 3), while the fourth was close to non-significance (Z score = 4.01) (Fig. 3a). By contrast, the remaining tests exhibited significantly positive *D* values. For the two parental species pairs with repeated hybrid lineages established, we found that each hybrid genome could resample separately different genomic components from their common parents, and novel genetic contents could occur through hybridization (Fig. 3c).

We subsequently identified genomic fragments sharing IBD for all diploid samples. We found nearly all large genomic fragments (mean 4.7 Mb, in contrast to the rest of mean size 150 kb) sharing IBD presented between diploid samples involved into hybridization (Fig. 3d). In particular, both *A. lijiangensis* and *A. hubeiensis* had 12-Mb IBD fragments, supporting the potential for replicate hybrid formation derived from a pair of common donor parents. A similar pattern between *A. fulvicoma* var. *fulvicoma* and *A. cylindrica* var. *reticulata* (*c.* 6 Mb) was observed. Using a kinship analysis (Manichaikul *et al.*, 2010), we characterized a duplicated/monozygotic (MZ) twin relationship for *A. lijiangensis* and *A. hubeiensis*, and a first-degree relationship for *A. fulvicoma* var. *fulvicoma* and *A. cylindrica* var. *reticulata* (Fig. 3d). These findings provided insights into the diversified patterns of kiwifruit hybridization, in terms of possible introgressive gene flow through rounds of backcrosses, leading to asymmetry in the tree support provided by AU tests and significant deviations of the *D* statistic (Fig. 3a) and also repeated hybrid speciation (Fig. 3c).

A dated backbone phylogeny

We finally reconstructed the phylogenetic tree by omitting lineages derived from reticulation as well as intraspecific polyploids. As expected, the resulting topologies of the TEN concatenated tree and the BCA species tree were completely identical (all 100% bootstrap support and high CF > 0.334) (Fig. 4a). The tree was characterized by an overall paraphyletic pattern, with the *A. arguta* clade at the most basal position. We refer to the tree as a 'backbone' phylogeny of kiwifruit plants as the taxa on it exhibited no evidence of hybrid origin but contributed to the reticulation of other taxa. This well-resolved backbone phylogeny allowed us to address the timing of kiwifruit diversification. For the three main *Actinidia* clades, we found that the *A. arguta* complex diverged from the ancestor of the other two clades *c.* 26.9 Ma, and then the *A. polygama* clade and HSF group diverged *c.* 18.9 Ma (Fig. 4a). The HSF group subsequently rapidly radiated into three clades of the *A. chinensis* complex, the *A. rufa* clade and the ancestor of the other HSF taxa *c.* 11 Ma. By contrast, most within-clade divergence occurred recently (*c.* 2.6–6.3 Ma) (Fig. 4a).

We investigated the genomic landscape of divergence between backbone taxa. We found that the pairwise divergence between them was significantly higher than that between taxa with parent−hybrid relationships (mean D_d value 0.84 vs 0.61, respectively; $P < 0.001$; *t*-test) (Fig. S17; Table S9). We detected an average of 3519 putative loss-of-function (LoF) mutations in protein-coding loci (Fig. S18; Table S10) for these backbone species and a predication of 279 species-specific novel genes by assembling their unmapped reads (Table S11). However, by searching core genomes (genes shared) that commonly occurred among backbone species, we found 11 982 genes (30.69% of the reference gene set) that were exactly identical in terms of the full length of the coding sequence regions (Fig. S19), suggesting extensive sorting of ancestral polymorphisms. We also found that the effective population size (N_e) in each backbone taxon has varied over time (Fig. 4b), with a notable decline in N_e among *A. chinensis* variety samples and *A. rufa* within the past 20 000–40 000 yr. We modeled changes of ecological niche distribution over time for seven backbone taxa with sufficient occurrence records (> 10 localities) (Notes S2). Notably, HSF taxa were possibly more sensitive to changes in ecological conditions than SSF taxa since the LGM (*c.* 0.022 Ma) (Fig. S20; Table S12).

Discussion

Whole-genome sequencing data are now widely used for investigating the evolutionary and demographic histories of both plants and animals (Martin *et al.*, 2013; Jarvis *et al.*, 2014; Novikova *et al.*, 2016). However, given the complicated backgrounds and diverse evolutionary processes of plant genomes, challenges are always there when we try to reconstruct the true species evolutionary network on the basis of genomic data alone (Posada, 2016). In the present study, we adopted the cautious approach of always using multiple analyses or lines of evidence to draw key conclusions. For example, we used both the concatenated method (the ML analysis) and the coalescent-based method (the BUCKy analysis) to infer the species phylogeny of *Actinidia* (Figs 1, S11). In the BUCKy analysis, we further estimated two species trees: the primary concordance species tree and the population tree. The primary concordance tree is a fully resolved tree but may include clades that are in < 50% of gene trees, while the estimated population tree is inferred from quartet CFs (Xin *et al.*, 2007). As for the results, the majority of these species trees are similar, but with alternative local branching patterns, suggesting reticulation,

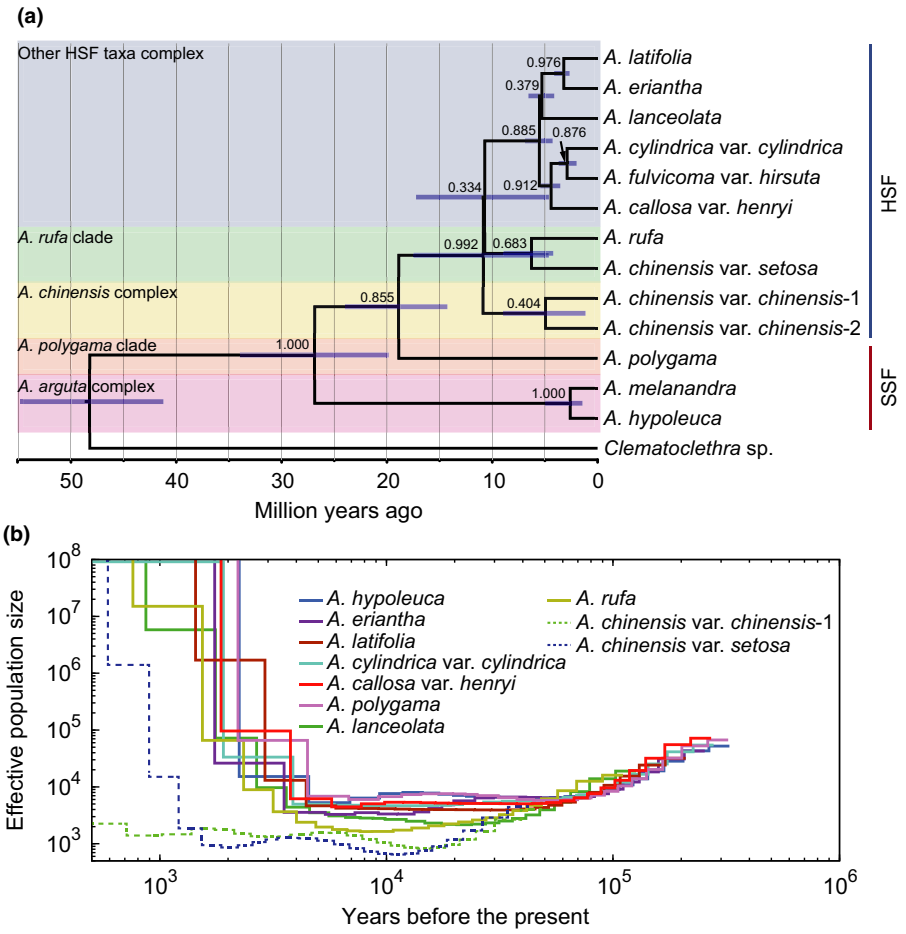

Fig. 4 Phylogeny and demography of backbone taxa. (a) The backbone tree consisting of diploid *Actinidia* taxa that exhibit no evidence of genomic admixture. All nodes are 100% bootstrap supported (not shown) and the primary Bayesian concordance factors (CFs) are given. Horizontal blue bars on each node indicate the 95% credible interval of divergence time in millions of years. SSF, smooth-skinned fruit; HSF, hairy and/or spotted (lenticillate) fruit. (b) Inferred ancestral effective population size (N_e) varies over time. We assumed a rate of 1×10^{-8} mutations per base pair per generation and a generation time of 7 yr for *Actinidia* under natural conditions.

rather than only a coalescent process occurring during the evolution of kiwifruit plants (Fig. S11). To determine a preferred genomic window size as a pseudo 'gene' for constructing gene trees, we repeated our analysis with variable genomic window sizes of 50, 100 and 200 kb, respectively, and found a nonsignificant effect of 'gene' size on the results in relation to the CFs and gene tree topographies (Notes S1). To our limited knowledge, there is not a fixed average size in relation to the (smallest) recombination unit of plant genomes. We therefore selected 100-kb nonoverlapping windows across the genomic alignment as 'genes'.

In addition to the nuclear trees constructed, we aligned reads to the respective cp and mt reference genomes to obtain additional information on variation and evidence supporting reticulation in the genus *Actinidia* (Fig. S13). Some recent studies also used this method to identify high-quality SNP variation in organellar genomes for phylogenetic reconstruction and detection of introgressive hybridization (Wu *et al.*, 2014). It should be noted that, for both cp and mt genomes, the method of assembly of each genome from each sample individually and subsequent alignment of these genomes to identify site variations is potentially more reliable. However, in our resequencing data, reads from some samples were not sufficient for assembly, and the majority of samples could only be assembled at the scaffold level. Comparatively, the mt genomes of kiwifruit are possibly more variable in terms of genomic structure and repeated sequence diversity, so a partial but not fully assembled reference genome can be constructed (Notes S1). Further investigations of both cp and mt genomes in *Actinidia* are thus needed in future studies.

Parallel to the challenges of using genomic data in investigations of plant reticulate evolution, the sampling strategy is also key to obtaining accurate results regarding hybridization and hybrid speciation between taxa within a hybrid swarm, and to accurately inferring their demographic history. Generally, using one individual of each taxon is not sufficient for performing both phylogenetic and demographic analyses, in particular for species with widespread distributions. To reduce this effect in our study, we increased representative samples for the two main widespread species complexes, *A. chinensis* and *A. arguta*, respectively (Notes S1; Table S1). Furthermore, we sequenced at least one sample in each taxon at a relatively high sequencing depth (e.g. $>20\times$; Table S2) to improve the reliability of the identification of SNP variation used in both phylogenetic and demographic analyses. According to data modeling carried out in previous studies, when sequencing depth is relatively high, demographic modeling using MSMC or PSMC approaches is generally reliable (Li & Durbin, 2011). Finally, to exclude the possibility that a sample was a hybrid individual derived from a hybrid zone, for sequencing analysis we selected only those taxa with clear taxonomic delimitations and an absence of morphological clines in previous

studies (Notes S1). Further population-level genetic analysis in relation to the hybrid taxa identified and, where present, their corresponding hybrid zones would be of value to dissect the eco-genetic mechanisms triggering hybrid speciation.

Using genome-wide data, we present here, for the first time, a clear picture of reticulate evolution in *Actinidia*. Clearly, *Actinidia* initially diversified via the rapid radiation of backbone lineages with adaptation to local environmental conditions (Fig. S21). These lineages then served as donor parents in the development of successive rounds of hybridization when demographic expansion led to range overlap, or unstable environmental conditions resulted in secondary contact (Fig. S22). As a representative genus among Chinese floras, we therefore argue here that the pattern identified in *Actinidia* may reflect a mode of plant diversification in this region, in terms of rapidly radiated lineages contributing to successive reticulation (Fig. 5).

In this mode of diversification, the role of the evolution of backbone taxa is critical for effectively maintaining and reinforcing the source of parental variation through rapid divergence. In our case, the divergence of backbone taxa was not a constant, gradual sequence of events. Instead, there were periodic bursts of rapid radiations, particularly for the HSF lineages, in which three main clades simultaneously diverged at *c.* 11 Ma, and further many within-clade divergences occurred recently (*c.* 2.6–6.3 Ma) (Fig. 4a), consistent with the extensive presence of ILS (Fig. S12). At deeper branches and timescales, the three main clades of the whole phylogeny – the ancestor of the HSF group, the *A. polygama* clade, and the *A. arguta* complex clade – revealed a pattern of paraphyletic evolution. Comparatively, we found higher numbers of LoF mutations in protein-coding loci in both the *A. polygama* clade and the *A. arguta* complex clade (with a representative taxon of *A. hypoleuca* investigated here) than in those within the HSF group (Table S10). This could reflect adaptation of clade-specific ecological preferences in relation to *de novo* mutation in genes associated with functional traits that differ between groups. Indeed, our ecological niche modeling showing that the SSF taxa were obviously more adapted to cold areas normally found in northern and western China than the HSF group (Fig. S20). However, our data cannot rule out the ILS

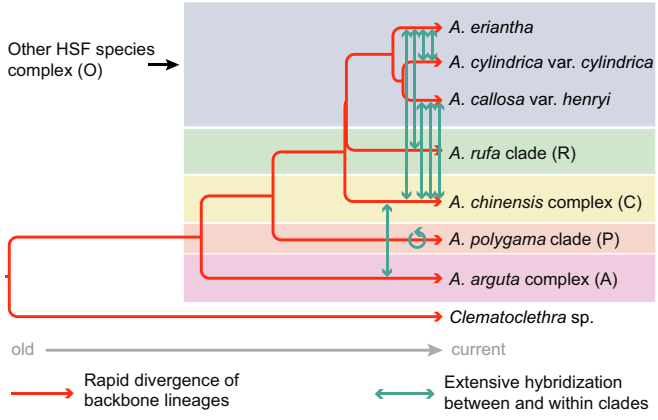

Fig. 5 A model of rapidly radiated backbone lineages contributing to successive reticulation. HSF, hairy and/or spotted (lenticillate) fruit.

presented in these deep branches completely. For example, both the consistently high CFs (0.203–0.401) which supported alternative gene trees at the corresponding nodes (Fig. S12b) and the overlapping in 95% credible intervals of divergence time (Fig. 4a) suggested shared ancestral variation among them, to a certain extent, during environment-specific sorting and divergence.

Through hybridization, the widespread establishment of hybrid swarms and syngameon phases is the second critical layer in this mode of diversification to maintain and prolong the momentum of radiation, which further increases diversity, if ecological conditions are also conducive to the establishment of hybrid populations. Surprisingly, although there is widespread establishment of allopolyploid hybrids in *Actinidia*, homoploid hybrid speciation is prevalent in this genus, particularly between closely related taxa within the HSF group (Fig. 3). In plants, *c.* 20 well-established homoploid hybrid species are known, but they are hard to detect (Mallet, 2007). Many known homoploid hybrid species are ecologically divergent from their parent species and often occur in more extreme environments than those of their parents, although cases exist where a homoploid hybrid species occurs in a habitat located between or close to those of its parent species (Sun *et al.*, 2014). The best documented examples are those of a study on the desert sunflower hybrid derivatives of two widespread species, *Helianthus annuus* and *Helianthus petiolaris* (Rieseberg *et al.*, 2003), and a recent study to test the hypothesis of a homoploid hybrid origin for the spruce species *Picea purpurea* (Sun *et al.*, 2014). Both intrinsic processes boosting genetic variance (including generating transgressive variation) and extrinsic opportunities in relation to unexploited ecological niches have been implicated in the origin and stabilization of homoploid hybrid taxa (Arnold, 2015).

Interestingly, most *Actinidia* homoploid hybrid taxa are very limited in geography in comparison to their parents (e.g. *A. lijiangensis*, *A. hubeiensis*, *A. cylindrica* var. *reticulata*, *Actinidia indochinensis*, and *A. zhejiangensis* in Fig. S2). Given the widespread occurrence and adaptation of their parental lineages with respect to geography, ecological niches left for hybrid taxa are possibly much more restricted in local areas. Alternatively, the establishment of a hybrid lineage could remain in its initial stage without widespread colonization and demographic expansion. Colonization of ecological niches would be a dynamic process, with competition between parental lineages and hybrid populations. The hybrid *A. fulvicoma* var. *fulvicoma* should be a winner in the competition for ecological niches because one of its parents, *A. cylindrica*, has a very restricted distribution within its wide geographic range (Fig. S2), suggesting that hybridization can widen ecological adaptation (Choler *et al.*, 2004). Although the parental lineages of two SSF hybrids could not be determined in our study (Notes S1), the widespread distribution, in particular for *A. kolomikta*, also revealed that ecological success may be attributed to a relatively old homoploid hybrid origin.

The two layers of rapid radiations in this mode of diversification are not independent; that is, following the continuous growth of the backbone tree, interspecific gene flow can have occurred historically between early-branching lineages and also recently between individual, newly established populations,

integrating into a phase of syngameon. The impact of recent changes in relation to geography and climate within and around China cannot be underestimated because of their critical role in providing ecological opportunities fueling rapid speciation (including widespread hybrid speciation). Under the action of the Himalayan movement and the integration of the Eurasian Continent, the three great topographic steps of major landforms in China have occurred. In particular, the recent uplift of the Qinghai-Xizang Plateau over the last 3.4 Myr significantly changed the geology and geomorphology of central and western China (Zheng, 2013). These processes collectively created huge mountain systems, which occupied *c.* 65% of the total land area of China, as well as thousands of islands in China's neighboring seas, resulting in extremely diverse habitats and ecological conditions that strongly and rapidly drove selection and local adaptation, leading to important effects on the demography and evolutionary patterns of taxa living within it.

Our data showed that the regional changes in both geography and climate had substantial impacts on the demography of *Actinidia* backbone taxa during rapid radiations (Fig. S20). Species with past population declines were those mainly distributed in eastern and southern China, including Taiwan and the southern parts of Korea and Japan (Fig. S2). Taking *A. chinensis* var. *setosa* as an example, the ancient geological origin of Taiwan island with recently active orogeny (Lin, 2002) could have increased the genetic isolation of this species from others distributed on the mainland, which in turn would have led to a decline in N_e (if widespread inbreeding occurred). A similar scenario is also possible for *A. rufa*, which mainly occurs in archipelagos along the Asian continental margin (Fig. 1d). Recent climatic changes in response to glaciation could also have driven changes in *Actinidia* demography, such as the limited distribution of the *A. chinensis* complex in southern and eastern refugia during the LGM (*c.* 0.022 Ma) (Fig. S20), which was consistent with the demographic decline of this complex which started between 20 000 and 40 000 yr ago (Fig. 4b). Furthermore, our ecological niche modeling revealed a dynamic distribution of different taxa during the Quaternary climatic oscillations (Fig. S20), which could play an important role in causing secondary contact, hybridization and speciation among previously diverged *Actinidia* taxa. Indeed, in many instances of plant hybrid speciation in China, such as the diploid hybrid origin of *Pinus densata* (Gao *et al.*, 2012), *P. purpurea* (Sun *et al.*, 2014) and *Ostryopsis intermedia* (Liu *et al.*, 2014), the Quaternary climate change and related geological events (e.g. the uplift of the Qinghai-Xizang Plateau) have been proposed to be the main triggers for the origin of new hybrid species.

In summary, the resolution afforded by data on genome-wide variation allowed us not only to resolve the complex hybrid relationship of kiwifruit plants, but also to refine a backbone phylogeny underlying reticulation. Our findings shed light on a special mode of rapid plant diversification in the face of diverse eco-environmental conditions. The notion that 'Web of Life' processes are fundamental to life on Earth is increasingly accepted. However, diversification modes within the web as a whole are heterogeneous depending on the organisms or eco-geographic regions examined. The two-layer mode found in kiwifruit is strongly associated with recent rapid changes in geography and climate within and around China and reflects the important roles of both intrinsic variable sources of genomic variability and extrinsic ecological opportunities. In particular, homoploid hybrid speciation represents an important mechanism fueling rapid reticulate radiations.

Interest in using wild relatives for crop breeding is currently increasing. Current commercial kiwifruit cultivation is almost entirely based on the *A. chinensis* complex, which has a short history of domestication since 1904 (Ferguson, 2004), and represents a very small fraction of the natural resources of *Actinidia*. Our whole-genome data for both hybrid and backbone lineages will also be valuable in guiding the classification (Notes S1), conservation and selection of natural resources for kiwifruit breeding applications. In view of the practice of large-scale monoculture in the global kiwifruit industry and the crisis of bacterial canker disease affecting cultivars, breeding approaches in which natural hybrids are directly improved or backbone taxa are introgressed into cultivars offer promising means of reinforcing the genetic basis of product diversity, including incorporation of tolerance and/or resistance to biotic and abiotic stresses under conditions of global change.

Acknowledgements

We thank Ying Zhang, Rui Wang, Lei Zhang, Yanchang Wang and Robert Schaffer for providing plant material and sequence data, Sheng Yu, Wenyan Fu, Ruibin Sun and Shaoxiong Gao for help with statistical analysis and validation experiments, and A. Ross Ferguson and Michael L. Arnold for reading and commenting on the manuscript. This work was supported by the Science and Technology Service Network Initiative of the Chinese Academy of Sciences (KFJ-EW-STS-076), the National Natural Science Foundation of China (31471847 and 31572092), and the Crop Germplasm Conservation Project of China (2016NWB027). The sequence data have been deposited in the NCBI Sequence Read Archive (SRA) under accession SRP073634.

Author contributions

Y.L., C.S., C.Z. and H.H. designed the project. D.L., Q.Z., X.Y., Z.W., Y.W. (Chinese Academy of Sciences, China), Y.G., S.W., X.L. and C.L. helped with the execution of the fieldwork and DNA extractions. X.Z., Y.W. (Wuhan Benagen Tech Solutions Company Limited, China), S.Z., L.L., H.C.M., W.H., Y.N., M.C. and L.D. performed most of the genome sequencing and bioinformatics analyses. J.G., P.M.D. and E.H. provided data or tools. Y.L., D.L., Q.Z. and H.H. wrote the manuscript.

References

Arnold ML. 2015. *Divergence with genetic exchange*. Oxford, UK: Oxford University Press.

Burke JM, Arnold ML. 2001. Genetics and the fitness of hybrids. *Annual Review of Genetics* **35**: 31–52.

Chat J, Jáuregui B, Petit RJ, Nadot S. 2004. Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany* **91**: 736–747.

Chat J, Petit RJ, Chalak L. 1999. Strict paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in intraspecific crosses of kiwifruit. *Theoretical and Applied Genetics* **99**: 314–322.

Choler P, Erschbamer B, Tribsch A, Gielly L, Taberlet P. 2004. Genetic introgression as a potential to widen a species' niche: insights from alpine *Carex curvula*. *Proceedings of the National Academy of Sciences, USA* **101**: 171–176.

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. 2015. Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. *Molecular Plant* **8**: 831–846.

Conesa A, Götz S, García-Gómez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

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

Ferguson AR. 2004. 1904 – the year that kiwifruit (*Actinidia deliciosa*) came to New Zealand. *New Zealand Journal of Crop and Horticultural Science* **32**: 3–27.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E *et al.* 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**: 1258524.

Gao J, Wang B, Mao J-F, Ingvarsson P, Zeng Q-Y, Wang X-R. 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Molecular Ecology* **21**: 4811–4827.

Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht MK, Macintyre RJ, Clegg MT, eds. *Evolutionary biology*. New York, NY, USA: Springer US, 93–120.

Gavrilets S, Losos JB. 2009. Adaptive radiation: contrasting theory with data. *Science* **323**: 732–737.

Grant V. 1981. *Plant speciation*. New York, NY, USA: Columbia University Press.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y *et al.* 2010. A draft sequence of the Neandertal genome. *Science* **328**: 710–722.

Hasumi H, Emori S. 2004. *K-1 coupled gcm (miroc) description*. Tokyo, Japan: Center for Climate System Research.

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**: 94–98.

Huang H. 2014. *The genus Actinidia, a world monograph*. Beijing, China: Science Press.

Huang S, Ding J, Deng D, Tang W, Sun H, Liu D, Zhang L, Niu X, Zhang X, Meng M *et al.* 2013. Draft genome of the kiwifruit *Actinidia chinensis*. *Nature Communications* **4**: 2640.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT *et al.* 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**: 1320–1331.

Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliussen TS, Vilstrup JT *et al.* 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences, USA* **111**: 18655–18660.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**: 371–375.

Larget BR, Kotha SK, Dewey CN, Ané C. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* **26**: 2910–2911.

Leducq J-B, Nielly-Thibault L, Charron G, Eberlein C, Verta J-P, Samani P, Sylvester K, Hittinger CT, Bell G, Landry CR. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1**: 15003–15010.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li J, Huang H, Sang T. 2002. Molecular phylogeny and infrageneric classification of *Actinidia* (Actinidiaceae). *Systematic Botany* **27**: 408–415.

Li JQ, Li XW, Soejarto DD. 2007. A revision of the genus *Actinidia* from China. *Acta Horticulturae* **753**: 41–44.

Lin CH. 2002. Active continental subduction and crustal exhumation: the Taiwan orogeny. *Terra Nova* **14**: 281–287.

Linder CR, Rieseberg LH. 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* **91**: 1700–1708.

Liu B, Abbott RJ, Lu Z, Tian B, Liu J. 2014. Diploid hybrid origin of *Ostryopsis intermedia* (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change. *Molecular Ecology* **23**: 3013–3027.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**: 18.

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* **46**: 523–536.

Mallet J. 2007. Hybrid speciation. *Nature* **446**: 279–283.

Mallet J, Besansky N, Hahn MW. 2016. How reticulated are species? *BioEssays* **38**: 140–149.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**: 2867–2873.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**: 1817–1828.

Mavárez J, Linares M. 2008. Homoploid hybrid speciation in animals. *Molecular Ecology* **17**: 4181–4185.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al.* 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.

Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends in Ecology & Evolution* **28**: 719–728.

Nolte AW, Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends in Genetics* **26**: 54–58.

Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Paape T, Schmid K, Fedorenko OM *et al.* 2016. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics* **48**: 1077–1082.

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

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.

Posada D. 2016. Phylogenomics for systematic biology. *Systematic Biology* **65**: 353–356.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* **317**: 910–914.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46**: 919–925.

Seehausen O. 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution* **19**: 198–207.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**: 492–508.

Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* **60**: 561–588.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**: 456–463.

Sun Y, Abbott RJ, Li L, Li L, Zou J, Liu J. 2014. Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Molecular Ecology* **23**: 343–359.

Testolin R, Cipriani G. 1997. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. *Theoretical and Applied Genetics* **94**: 897–903.

Whitfield JB, Lockhart PJ. 2007. Deciphering ancient rapid radiations. *Trends in Ecology & Evolution* **22**: 258–265.

Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, Perrier X, Ruiz M, Scalabrin S, Terol J *et al.* 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology* **32**: 656–662.

Xin L, Ma B, Zhang K. 2007. A new quartet approach for reconstructing phylogenetic trees: quartet joining method. In: Lin G, ed. *Computing and combinatorics. Lecture notes in computer science*. Berlin/Heidelberg, Germany: Springer, 40–50.

Yakimowski SB, Rieseberg LH. 2014. The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *American Journal of Botany* **101**: 1247–1258.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. 2015. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS ONE* **10**: e0129347.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.

Zheng D. 2013. Physical geography. In: Hong D, Blackmore S, eds. *Plants of China*. Cambridge, UK: Cambridge University Press, 35–54.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Fruit, flower and leaf morphology of the *Actinidia* taxa sequenced.

Fig. S2 The approximate geographic distribution of the *Actinidia* taxa sequenced.

Fig. S3 Bases mapped as a function of sequencing depth for *Actinidia* samples.

Fig. S4 Proportion of missing data as a function of the sequencing depth for *Actinidia* samples.

Fig. S5 Pipeline of mitochondrial genome assembly.

Fig. S6 The relative heterozygosity distribution of tetraploid and hexaploid samples.

Fig. S7 Chromosomal distributions of bi-, tri- and tetra-allelic SNPs in two representative tetraploid and hexaploid samples of *Actinidia chinensis*.

Fig. S8 Tree topologies employed to investigate parental genomic contributions in the AU test.

Fig. S9 Total evidence nuclear tree of 40 sequenced *Actinidia* samples.

Fig. S10 Primary concordance species trees produced from both 50-kb and 200-kb genomic windows.

Fig. S11 Incongruence between alternative species trees.

Fig. S12 Anticipated incomplete lineage sorting (ILS) identified in the BUCKy analysis.

Fig. S13 Cytonuclear conflicts among different *Actinidia* lineages.

Fig. S14 Genome-wide heterozygosity of all diploid samples.

Fig. S15 Asymmetric discordances identified between single taxon pairs.

Fig. S16 Parental genomic contributions revealed by 10-kb nonoverlapping windows along *Actinidia* chromosomes.

Fig. S17 Pairwise genetic divergence of diploid *Actinidia* samples.

Fig. S18 GO enrichments of genes with loss-of-functions (LoF) variants in backbone species.

Fig. S19 Gene ontology (GO) annotations on genes shared among backbone kiwifruit species.

Fig. S20 Ecological niche modeling of *Actinidia* backbone taxa.

Fig. S21 Rapid local ecological adaptation of *Actinidia* backbone clades over time.

Fig. S22 *Actinidia* backbone taxa serving as the source of evolutionary variation to contribute to repeated and continuous hybrid formation.

Table S1 List of sequenced *Actinidia* samples

Table S2 Sequencing summary of all *Actinidia* samples

Table S3 Distribution of SNP variants along chromosomes

Table S4 Reads mapped to the chloroplast reference genome

Table S5 Reads mapped to the mitochondrial reference genome

Table S6 Genome-wide heterozygosity of diploid *Actinidia* samples

Table S7 Proportion of the gene trees supporting alternative topologies estimated by AU tests and Patterson's *D* statistic calculated for taxa with significant BUCKy discordance (CF ≥ 0.05)

Table S8 Parental genomic contributions estimated by AU tests and Patterson's *D* statistic

Table S9 The mean pairwise divergence (D_d values) between *Actinidia* diploid samples

Table S10 Variation consequences of coding sequence variant of backbone taxa

Table S11 Genes predicted based on *de novo* assembly of unmapped reads in backbone taxa

Table S12 The bioclimatic variables extracted from the WorldClim database and their effects on the distribution of *Actinidia* backbone taxa

Notes S1 Supplementary notes.

Notes S2 SNPs for experimental validation, concordance factors of BUCKy analysis, Patterson's *D* statistic calculated for each chromosome, gene ontology annotations of predicted genes and geographic occurrence records of *Actinidia* backbone taxa.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.