Untying Gordian knots: unraveling reticulate polyploid plant evolution by genomic data using the large Ranunculus auricomus species complex

Kevin Karbstein1,2, Salvatore Tomasello1, Ladislav Hodač1,3, Natascha Wagner1, Pia Marincek1,5, Birthe Hilkka Barke1, Claudia Paetzold1,4, and Elvira Horandl1

1Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), Albrecht-von-Haller Institute for Plant Sciences, University of Göttingen, 37073 Göttingen, Germany; 2Georg-August University School of Science (GAUSS), University of Göttingen, 37073 Göttingen, Germany; 3Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany; 4Department of Botany and Molecular Evolution, Senckenberg Research Institute, 60325 Frankfurt (Main), Germany

Abstract

- Speciation via hybridization and polyploidization is a major evolutionary force in plant evolution but is still poorly understood for neopolyploid groups. Challenges are attributed to high heterozygosity, low genetic divergence, and missing information on progenitors, ploidy, and reproduction. We study the large Eurasian Ranunculus auricomus species complex and use a comprehensive workflow integrating reduced-representation sequencing (RRS) genomic data to unravel reticulate evolution, genome diversity and composition of polyploids.
- We rely on 97,312 restriction site-associated DNA sequencing (RAD-Seq) loci, 576 targeted nuclear genes (48 phased), and 71 plastid regions derived from 78 polyploid apomictic taxa and four diploid and one tetraploid putative sexual progenitor species. We applied (phylo)genomic structure, network, and single nucleotide polymorphism (SNP)-origin analyses.
- Results consistently showed only 3–5 supported and geographically structured polyploid genetic groups, each containing extant sexual and one unknown progenitor species. Combined analyses demonstrated predominantly allopolyplod origins, each involving 2–3 different diploid sexual progenitor species. Young allotetraploids were characterized by subgenome dominance and nonhybrid SNPs, suggesting substantial post-origin but little lineage-specific evolution.
- The biodiversity of neopolyploid complexes can result from multiple hybrid origins involving different progenitors and substantial post-origin evolution (e.g. homoeologous exchanges, hybrid segregation, gene flow). Reduced-representation sequencing genomic data including multi-approach information is efficient to delimit shallow reticulate relationships.

Introduction

Polyploidy, the presence of more than two chromosome sets, occurs across eukaryotes (Otto & Whitton, 2000; van de Peer et al., 2017). All flowering plants are of ancient polyploid origin, and several additional polyploidization events were found in various lineages (van de Peer et al., 2017; Leebens-Mack et al., 2019). Young polyploidization events (neopolyploidy) tend to be followed by an upshift of diversification rates (Landis et al., 2018). Two different main polyploidization types exist: allopolyploids are formed by hybridization between different species/lineages whereas autopolyploids arise within species (Comai, 2005; Blischak et al., 2018). Consequently, autopolyploids contain genetically similar subgenomes whereas allopolyploids are composed of previously diverged subgenomes. Both types can occur within the same species or species complex (Horandl, 2022). Allopolyploidy is frequently associated with higher degrees of genomic, transcriptomic, and epigenomic changes than autopolyploidy, and is thus considered particularly likely to create novel genomic features (Abbott et al., 2013; Wendel, 2015; Spoelhof et al., 2017; Rothfels, 2021).

Polyploid lineages are not only shaped by evolutionary origin and genomic contributions of progenitors, but also by post-origin processes, resulting in a mosaic-like genome structure (Soltis et al., 2015). Expression bias due to epigenetic changes and homoeologous exchanges (HEs) directly after allopolyploidization can distort original genomic contributions, leading to subgenome expression and/or sequence dominance (Blischak et al., 2018; Alger & Edger, 2020). Moreover, Mendelian segregation in the first diploid hybrid generations before polyploidization, backcrossing of polyploids to their sympatric progenitors, and/or gene flow among polyploid lineages might influence original contributions (Hodač et al., 2018, 2019; Melichárková et al., 2020; Wagner et al., 2020).
In plants, polyploidization and/or hybridization are frequently connected to apomixis, i.e. reproduction via asexually-formed seeds (Askèr & Jerling, 1992; Hojsgaard & Hörandl, 2019). Apomixis is usually facultative, and after multiple hybrid origins, residual sexuality allows for backcrossing to progenitors and intercrossing of polyploids, resulting in large networks of numerous hybridogenetic lineages (Fig. 1). With increasing ploidy and/or obligate apomixis, these lineages are expected to become invariable due to loss of recombination and gene flow with other lineages (Grant, 1981; Coyne & Orr, 2004; Hörandl, 2022; Fig. 1). Allelic sequence divergence (Meselson effect) can further augment heterozygosity (Welch & Meselson, 2000; Pellino et al., 2013; Hodač et al., 2019). Heterozygosity has several benefits, for example, increased genetic flexibility, buffering deleterious mutations, or changes in secondary metabolites, probably leading to the ability of polyploids in settling more climatically variable habitats (Rice et al., 2019; Qiu et al., 2020; Karbstein et al., 2021a).

Plant species complexes link microevolutionary processes (polyploidization and hybridization) with macroevolutionary patterns (speciation), and are thus appropriate for understanding young evolutionary processes (Soltis et al., 2015; Pinheiro et al., 2018). The sexual progenitors of such complexes are already characterized by low genetic divergence, incomplete lineage sorting (ILS), and/or partial hybridogenic origins (Hörandl, 2018; Wagner et al., 2019; Karbstein et al., 2020c). Traditional markers from single organellar or few nuclear DNA regions (Freeland et al., 2011; Karbstein et al., 2019, 2020a) often failed to reconstruct reticulate relationships in polyploid complexes even at the diploid level due to uniparental inheritance, marker-specific evolution, and/or low variability (Kirschner et al., 2015; Fehr et al., 2021; Rothfels, 2021). Consequently, polyploids have been frequently avoided in phylogenetic studies (Rothfels, 2021).

Reduced representation sequencing (RRS; Davey et al., 2011) genomic data provide orders of magnitude more information than traditional markers and have proven to be effective at resolving diploid phylogenetic relationships among species that diversified c. 0.1–50 million years ago (Ma) (e.g. Pellino et al., 2013; Hipp et al., 2014; Eaton et al., 2017; Carter et al., 2019; Wagner et al., 2020). Restriction site-associated DNA sequencing (RAD-Seq) covers a subset of noncoding and coding regions across the entire genome and is mostly used for analyzing closely related groups within species or genera (Davey et al., 2011; McKain et al., 2018). Target enrichment aims at collecting rather conservative low-copy nuclear genes able to resolve relationships at the genus level (Schmickl et al., 2016; Carter et al., 2019). Although RAD-Seq yields many more loci and single nucleotide polymorphisms (SNPs) than target enrichment, read assembly is bioinformatically more challenging and locus dropout can become problematic with increasing genetic divergence (Eaton et al., 2017; Karbstein et al., 2020c). Target enrichment loci are usually longer, allowing for allele phasing, gene tree estimation, and coalescent-based approaches. Allelic information is particularly important for correct phylogenetic inferences in highly reticulate, young evolutionary relationships (Eriksson et al., 2018). In addition, while sequencing targeted nuclear genes, plastid data can be easily gained from off-target reads (Hyb-Seq) (Weitemier et al., 2022).

---

**Fig. 1** Evolutionary processes in neopolyploid species complexes and methods to address these processes. The figure was drawn to trace evolutionary processes in young species groups, like the here studied *Ranunculus auricomus* complex. (a) Evolution of an apomictic polyploid complex from two sexual progenitor species and evolution of lineages after origin (concept of Babcock & Stebbins (1938); see also Grant (1981) and Coyne & Orr (2004) for modern interpretations). (b) Description of evolutionary processes and applied methods. CP, chloroplast (plastid) regions; RAD-Seq, restriction site-associated DNA sequencing loci; TEG, target enriched nuclear genes. See Fig. 2 for the sampling and Fig. 3 for the detailed workflow applied in this study.
et al., 2014; Folk et al., 2015). Nuclear-plastid discordances help to identify past hybridization or allopolyploidization events (Huang et al., 2014; Stull et al., 2020) and the maternal progenitor of polyploids. Consequently, a promising approach is the combination of both RAD- and Hyb-Seq to cover the entire plant genome and unravel reticulate polyploid plant evolution.

Reticulate relationships should be inferred by nonbifurcating network-like algorithms, as bifurcating models lead to incongruences in tree reconstructions (McDade, 1992, 1995; Rothfels, 2021). Recently developed software can (1) unravel genetic structure and composition (e.g. STACEY: Jones, 2017b; RADAISER: Malinsky et al., 2018); (2) model network-like evolution under the coalescent model (e.g. PHYLONETWORKS: Solis-Lemus et al., 2017; PHYLONET: Than et al., 2008; Wen et al., 2018); or (3) infer polyploid genome evolution (e.g. SNIPLOID: Peralta, 2018); or (3) infer polyploid genome evolution (e.g. RADPAINTER: Malinsky et al., 2018); or (3) infer polyploid genome evolution? (1) Do RAD-Seq and phased nuclear gene data reflect a clear genetic and/or geographical structure? (2) Do nuclear and plastome data deliver conflicting signals? (3) Are ampicic lineages of autopolyploid or allopolyploid origin? (4) How many progenitors contributed to their genomes? (5) To which extent are polyploid genomes influenced by post-origin evolution?

Materials and Methods

Population sampling

We included all five sexual species and 78 of the most widespread, triploid to hexaploid ampicic *R. auricomus* taxa (Fig. 2, see references in legend). We collected 293 samples from 248 populations across Europe for genetic analyses (Supporting Information Table S1). The sampling included three populations per taxon on average, and 29 diploid sexual, three triploid, 206 tetraploid, and 10 hexaploid facultative to obligate ampicic populations (Karbstein et al., 2021a). Further details about locations, ploidy, reproduction modes, samples per population, and material for DNA sequencing are given in Table S1. In almost all cases, we selected the same samples among datasets. Data sampling and the workflow are described in Fig. 3.

Laboratory work, locus assembly, and parameter optimization

We used the already sequenced 280 *R. auricomus* and two outgroup RAD-Seq samples from Karbstein et al. (2020c, 2021a). DNA extraction, quality check, RAD library preparation, single-end sequencing of 100 bp reads (Baird et al., 2008), and subsequent quality filtering also followed Karbstein et al. (2020c, 2021a). For target enrichment, we added 85 newly sequenced samples to the already existing 28 samples sequenced by Tomasello et al. (2020), totaling 113 accessions. All plastid data (CP) from off-target reads is published herein. The employed bait set, library preparation, hybrid capture, and Illumina sequencing followed Tomasello et al. (2020).

For de novo assembly of RAD-Seq loci and parameter optimization, we used ipyrad v.0.9.14 (Eaton & Overcast, 2020). We exactly followed Karbstein et al. (2021a): the within-sample clustering was optimized separately for each ploidy level balancing number of clusters, cluster depth, and clusters rejected due to high heterozygosity. Then, the among-sample clustering (merged assembly) was optimized for number of polymorphic loci, SNPs, loci filtered by maximum number of SNPs, removed duplicates, shared loci, and new polymorphic loci. To assess the effects of number of loci and missing data on (phyl)genomic analyses, we selected different minimum amounts of samples per locus, i.e. 74% (‘min10’), 55% (‘min30’), and 44% (‘min50’) missing data. We allowed a maximum number of four alleles per locus because the majority of samples were tetraploid and alignments showed almost never more than four alleles (randomly checked with
GENEIOUS v.R11 2020.2.4, Kearse et al., 2012). Moreover, diploid base calling of IPYRAD leads to lumping of alleles in only two phases/haplotypes. Nevertheless, loss of genetic information is unlikely due to rare occurrence of more than two bases per site (randomly checked with GENEIOUS) and low overall genetic differentiation. In addition, choosing only one RAD-Seq SNP per locus (SNMF, PHYLOGENETS) eliminates the bias of merging different polyploid subgenomes in genetic analysis.

For target enrichment data analysis, reads were processed with HYBPHYLOMAKER v.1.6.4 (Fér & Schmickl, 2018), using target regions as pseudoreference for read mapping (Notes S1; Table S2). Data filtering yielded 579 genes from which the most informative, nonhomoplasious, and free-from-paralog-sequences 48 genes were selected (Notes S2; Table S3). We phased the selected loci using a similar approach as described in Eriksson et al., 2018 (Table S4). We processed the mapped BAM files of all samples with SAMTOOLS v.0.1.19 (Li et al., 2009). The polyploid samples were phased further, looking at the phased BAM files in IGV v.2.8.9 (Robinson et al., 2011) and manually adding additional alleles to the alignments when detected. Off-target plastid reads were also assembled with HYBPHYLOMAKER using the Ranunculus repens plastome (Dann et al., 2017; Table S5) as reference, excluding plastome regions with high amounts of missing data (c. 50% plastome completeness; Notes S3; Table S6).

Genetic structure (RAD-Seq)
To investigate genetic structure, we first conducted analyses with RADPAINTER+FINERADSTRUCTURE v.0.3.2 (Malinsky et al., 2018) using the alleles.loci IPYRAD files (all SNPs per locus). RADPAINTER calculates the nearest neighbor haplotype coancestry values, which gives information about genetic similarity of an individual to all other individuals across alleles and loci. We used FINERADSTRUCTURE to assign individuals to groups (1000 000 burn-in; 1000 000 iterations) including a tree building MCMC algorithm (100 000 burn-in), and plotted the results using ‘FINERADSTRUCTUREPLOT.R’ (R v.4.0.3; R Core Team, 2020). We selected the min10 dataset (minimal 10% samples per locus; 97 312 loci, 74% missing data) for further interpretations because it yielded the best genetic resolution (Fig. S1).

Moreover, we performed genetic structure analyses applying SNMF within the R package ‘LEA’ v.3.0.0, which handles mixed-ploidy datasets, does not rely on assumptions of Hardy–Weinberg equilibrium, and is thus particularly suitable for analyzing polyploid apomicts (Frichot et al., 2014; Frichot & François, 2015; Weiss et al., 2018). We used ugeno IPYRAD files (one SNP per locus), and set the number of genetic clusters (K) from 1 to 80, maximal ploidy to 4, and repetitions to 7. To choose the number of ancestral K values, we used the implemented cross-entropy criterion. We found the optimal K values between 3 and 5 (Figs S2–S8). Maps were drawn using POPSUTILITIES.R (http://membrres-timc.imag.fr/Olivier.Francois/POPSutilities.R). We selected the min30 dataset (minimal 30% samples per locus; 33 165 loci, 55% missing data), which balanced number of loci and missing data (Fig. S4).

Genetic structure (nuclear genes)
To unravel genetic structure based on phased nuclear genes, we utilized the species delimitation software STACEY v.1.2.1...
This coalescent-based approach models the genetic structure using allelic information (Jones, 2017b; Andermann et al., 2019), and is appropriate for mixed-ploidy datasets. Input files were prepared in BEAUTI v.2.6.1 (Bouckaert et al., 2014) using the 48 phased loci. Each sample was treated as ‘minimal cluster’ (alleles of one accession are not allowed to be split into different species). Sequence substitution models were selected for each locus separately using the Bayesian information criterion (BIC) in MODELTEST-NG v.0.1.6 (Darriba et al., 2020). Substitution models, clock models, and gene trees were treated as unlinked for all loci. To reduce the search space, parameters of the substitution models were fixed to those found in MODELTEST-NG. Detailed STACEY settings are described in Notes S4. The analyses were run for $2 \times 10^9$ iterations sampling every 200,000th generation in BEAST v.2.6.1 (Bouckaert et al., 2014). Four independent runs were performed and, after checking convergence between independent analyses and effective samples size values (ESS > 200, minority of parameters 100–200) in TRACER v.1.6 (Rambaut et al., 2018), we combined tree output files using LOGCOMBINER v.2.6.1 (Bouckaert et al., 2014) and discarding 10% of the analyses as burn-in. The obtained file was processed with the ‘SpeciesDelimitationAnalyser’ (Jones et al., 2015). The similarity matrix was produced using the R script of Jones et al. (2015).

Plastome (CP) and nuclear (RAD-Seq, TEG) phylogenies

To investigate the presence of nuclear-plastid discordances, we used 71 selected plastid regions to infer a maximum likelihood (ML) tree and 100 Bootstrap (BT) replicates with RAxML-NG v.0.9.0 (Kozlov et al., 2019). Models of sequence evolution were assessed for each region separately using MODELTEST-NG. All alignments were concatenated with AMAS v.0.98 (Borowiec, 2016) and different regions were treated as different partitions, each with its respective sequence evolution model. To gain additional information about haplotype evolution, we calculated neighbor-net networks running SPLITSTREE v.4.14.6 (Huson & Bryant, 2006) as described in Karbstein et al. (2021a).

For each of the nuclear datasets (RAD-Seq, TEG), trees and quartet sampling (QS) analyses (Pease et al., 2018) were calculated following Karbstein et al. (2021b) (Figs S11–S15). Since QS revealed several conflicting patterns, the assumptions of bifurcating relationships were clearly rejected, and hence trees were not used for further analyses.

Subgenome contribution of polyploids (RAD-Seq, nuclear genes)

To investigate polyploid genomes in more detail, we selected polyploid individuals with obvious reticulation signals per main
Table 1 Consensus results, plastid (CP) data, subgenome contributions, and formation types of tested tetraploid *Ranunculus auricomus* accessions (H₁–H₁₀).

| Analysis | H₁ 'R. × platycarpoides' | H₂ 'R. × elatior' | H₃ 'R. × pseudocassubicus' | H₄ 'R. × hungaricus' | H₅ 'R. × fissifolius' |
|----------|--------------------------|-------------------|---------------------------|----------------------|----------------------|
| Consensus results | C | N | F | E | F | E | F | N | C | F | N | E | E | N | F | N |
| Likel + AIC (PHYLONET) | Reticulate (N, C) | Reticulate (C, F) | Reticulate (C = E) | Reticulate (F, C > E) | Reticulate (3x E, F) |
| CP type | C/F | C/F | C/F | C/F | C/F |
| Final results | allo | allo | allo | allo | allo |
| Genome evolution | allo | allo | allo | allo | allo |
| No. sub-genomes/s | 2 | 2 | 2 | 2 | 3 |

Subgenome contributions/plastid types of polyploids from sexual progenitors are colored according to Fig. 2. Consensus results are based on restriction site-associated DNA sequencing (RAD-Seq; RADPAINTER+RADSTRUCTURE, PHYLONETWORKS) and phased nuclear target enrichment gene (TEG; STACEY, PHYLONETWORKS, PHYLONET) analyses. Final results indicate final subgenome contribution(s), i.e. consensus results corrected by the full likelihood approach + AIC calculations in PHYLONET (likel + AIC, AIC = Akaike information criterion; more than one result if AIC network difference was less than 10 units) and CP results (haplotypes; C/F = plastid type shared by the diploid sexual species *Ranunculus cassubicifolius* and *Ranunculus flabelliformis*, *not the same sample between CP and network analyses, # = haplotype from an unknown/extinct sexual progenitor species related to *Ranunculus envalirensis* of central Europe), genome evolution (allo, allopolyploid; auto, autopoly- ploid), and number of involved subgenomes. See also Figs 4–8, Supporting Information Tables S1, S7, and data on FigShare.
Phylogenetic network analyses (RAD-Seq)

To corroborate the already gained genetic structure information, we carried out analyses with maximum pseudolikelihood, coalescent-based PHYLONETWORKS v.0.12.0 using unlinked SNPs per species (Solís-Lemus et al., 2017). We applied the R function SNPS2CF.R v.1.2 (Olave & Meyer, 2020) to transform SNP-based min30 RAD-Seq alignments into quartets and quartet concordance factors (CF; proportion of locus-based trees supporting the quartet). A custom R script converted the ustr (one SNP per locus, two phases) IntrAD output file into an adequate input format for SNPS2CF. We created 10 subsets each containing one tetraploid accession (H₁–H₁₀) and all available accessions of diploid sexual progenitors (except *R. marsicus* due to no significant subgenome contribution in previous analyses). We specified ‘between species only’, no maximum number of SNPs, maximum number of quartets of 1000, and 100 BT gene trees were inferred with RAXML-NG for branch lengths and inheritance probabilities under full likelihood and rooted networks 

Phylogenetic network analyses (nuclear genes)

Moreover, we performed network analyses based on the 48 phased nuclear genes. We also investigated H₁–H₁₀, and used gene trees as input and two different maximum pseudolikelihood, coalescent-based approaches taking alleles per species into account: SNaQ implemented in PHYLONETWORKS as for RAD-Seq data and InferNetwork_MPL implemented in PHYLONET v.3.8.2 (Than et al., 2008; Wen et al., 2018). For each polyploid tested, alignments were modified to include all appropriate diploid accessions and the respective polyploid individual. Models of sequence evolution were selected with MODELTEST-diploid accessions and the respective polyploid individual. The same procedure was also applied to the STACEY posterior probability. To ensure comparability among datasets, we aimed at selecting the same individuals (except for the monophyletic taxon ‘*R. × elatior*’ (H₂), see also Figs S11–S15).

Origin of polyploids (RAD-Seq, nuclear genes, CP)

To infer parental subgenome contribution per polyploid, we listed all previously generated results in Table S7 (RAD-Seq, nuclear genes). Parent 1 (P₁) is always the parent with the largest subgenome contribution (coancestry/posterior probability values, inheritance probabilities) followed by the other parental contributions (P₂–P₄). We applied multiple criteria for building a consensus based on all generated results (Table S7): (1) Take the most abundant parent within a column; (2) if there are two equal abundant parents (e.g. two-times ‘C’ and ‘F’) within a column, both parental subgenome contributions were taken (‘C/F’). To validate the obtained consensus and to infer genome evolution (autopolyploid vs allopolyploid), we submitted all previous results before consensus building to the full likelihood approach implemented in PHYLONET (Yu et al., 2012). The CalcProb function calculates the likelihood of gene trees under a given species network and thus the total likelihood of the same network.

To include genetic structure results, networks were manually constructed using the tree backbone topology in Karbstein et al. (2020c) and the first two putative progenitors identified by these methods. The autopolyploid scenario was tested utilizing the ASTRAL III trees already used as starting tree for the PHYLONETWORKS analyses. We rooted all networks with *R. cassubicifolius* to make scenarios more comparable. To compare autopolyploid scenarios with allopolyploid ones, we scored results applying the Akaike information criterion (AIC), taking into account that the number of parameters in a tree/network is equal to the number of branch lengths plus (for the networks) the parental contributions (i.e. *k* = 8 and *k* = 13 for the tree and the networks, respectively). We determined the final subgenome contribution(s) by correcting the consensus results by the previously generated full likelihood approach results and inferred haplotypes (Table S7).

Origin of SNPs (RAD-Seq)

To investigate post-origin evolution of allopolyploids in more detail, we carried out SNPiloid (v.March 17, 2016; Peralta et al., 2013; Wagner et al., 2020; scripts on Github). SNPiloid compares an allotetraploid and a diploid putative parental species (DIPLOID2) with a diploid parental reference (DIPLOID1). The resulting SNPs were categorized: cat1&2 result from interspecific, post-origin hybridization (SNPs either DIPLOID1/2); cat3/4 represent lineage-specific, post-origin SNPs (SNP does not match SNPs of DIPLOID1/2); cat5 represents the homeo- or hybrid-SNPs from both parents from the allopolyploidization
event (Peralta et al., 2013; Wagner et al., 2020). For example, a first-generation hybrid is expected to have only homo-SNPs inherited from the parental species (cat5), and no interspecific SNPs (cat1&2) or derived SNPs (cat3/4).

We created references of diploids by merging all accessions of a single progenitor species into a single FASTQ file to include dominant parental SNPs and to reduce bias by selecting only one individual, and conducted within-sample clustering in IPYRAD.
(settings identical to Karbstein et al., 2020c). Obtained consensus files containing ambiguity codes (ignored) were used as DIPLOID1 (reference) and merged FASTQ files as DIPLOID2. We specified a minimum read depth per position of 20 to filter sequencing errors. We excluded the category ‘others’ (heterozygous positions of DIPLOID2) from final results. Moreover, we always observed dominance of interspecific SNPs of cat2 compared with cat1 SNPs, independent of parental combination. This was probably due to the majority rule base call references neglecting natural genetic variation. Therefore, we generally summarized both categories to ‘cat1&2’ to avoid biases within interspecific SNP category.

Results

Genetic structure analyses

RADSequencing+fineRADSTRUCTURE structure analysis based on RAD-Seq-SNPs revealed three supported main clusters (Fig. 4a). Sexual species clustered with several polyploid apomicts: (I) *R. cassubicifolius* (subgenome C) with tetraploid to hexaploid taxa, (II) *R. flabellifolius*, *R. marsicus*, and *R. notabilis* (subgenomes F, M, and N) with triploid to hexaploid taxa, and (III) *R. envalirensis* (subgenome E) with tetraploid taxa. Commonly, polyploids showed high coancestry values, i.e. orange to red colors, with different clusters indicating reticulation events (see particularly polyploids H1–H10; Fig. 4a). Highest values were found with sexual subgenomes occurring in the same cluster (Table S7). Polyploids of cluster I showed highest similarity values with subgenome C and lowest ones with subgenomes N and F. In contrast, polyploids of cluster II are genetically more heterogeneous and divided into subclusters Ia and IIb: polyploids of Ia shared high similarity values with subgenome N and low coancestry values with subgenomes F and C whereas polyploids of IIb showed low values with F and E. In cluster III, polyploids only exhibited high similarity to subgenome E.

STACEY analysis based on phased nuclear genes revealed similar results (Fig. 4b; Table S7). Each sexual species is surrounded by polyploid apomictic taxa, which showed several reticulations and highest posterior probabilities with intra-cluster sexual progenitor subgenomes. The former RAD-Seq cluster II is divided into three distinct clusters each containing a single sexual species (II–IV), and many polyploids of the former RAD-Seq cluster II are incorporated into cluster V. In addition, polyploids of cluster I also shared significant posterior probabilities to subgenomes E and polyploids of cluster V to subgenomes F, C, and N. The polyploid sexual subgenome M shared no significant coancestry/posterior probability values with polyploid apomicts. Whereas RAD-Seq clusters I/nuclear gene clusters I + II were predominantly characterized by undivided basal leaf taxa, the remaining clusters exhibited only dissected ones.

SNMF structure analysis based on unlinked RAD-Seq SNPs also unraveled three to four main clusters (Fig. 5a,b). Although polyploid apomictic taxa were characterized by a dominant genetic partition, they also showed 1–3 minor ones (Fig. 5a,b). The likeliest number of clusters, \(K = 3\), showed a west–east distribution of clusters across Europe (Fig. 5a,b). The clusters themselves are north–south distributed. *Ranunculus envalirensis* and related polyploids (E, green partition) mainly inhabit regions in western Europe. *Ranunculus flabellifolius*, *R. marsicus*, and *R. notabilis* and related polyploids (F, M, and N, orange partition) predominantly occupy central Europe. *Ranunculus cassubicifolius* and related polyploids (C, blue partition) are mainly distributed in central-eastern Europe. Results of \(K = 4\) showed the emergence of a central European cluster without a sexual species (gray partition) out of the former green one (Fig. 5a,b). The SNMF results are comparable to previous ones (except that the gray partition is predominantly found in RAD-Seq cluster III/nuclear gene cluster V).

Plastome phylogeny compared with nuclear data

The ML tree based on plastid regions (CP) revealed four well-supported main clades (i.e. haplotype groups; BT = 100; Figs 6a,
In general, within-clade relationships were mainly low or not supported (FDP < 70). Clade I consists of haplotypes from *R. cassubicifolius* (C), *R. flabellifolius* (F), and various related polyploids. Accessions of *R. cassubicifolius* and *R. flabellifolius* were completely intermingled, contrary to nuclear datasets (Figs 4, 5). Clade II contained only haplotypes from polyploid taxa. The remnant two haplotype clades III and IV consisted of *R. envalirensis* (E) and few polyploids, and *R. notabilis* (N), *R. marsicus* (M), and various polyploids, respectively. Interestingly, the diploid *R. notabilis* and the tetraploid sexual *R. marsicus* belong to the same haplotype group contrary to nuclear gene data (Fig. 4b). The splits graph of the neighbor-net analysis also exhibited four differentiated clusters with low genetic distance (Figs 6b, S10).

**Phylogenetic networks combined with genetic structure and plastid data**

Phylogenetic networks based on RAD-Seq-SNPs and alleles of phased nuclear genes mainly showed two different subgenome contributions for polyploids (Figs 7, 8; Table 1). These polyploids were usually characterized by a dominant (P1 = 51–99% inheritance probability, mean 74%) and a minor subgenome DNA sequence contribution (P2 = 1–49% inheritance probability, mean 26%). Concerning PHYLONET likelihood + AIC calculations, reticulate evolution and thus allopolyploid origin was confirmed in most cases. Within cluster I (Fig. 4a, b), polyploids H1–H4 possessed the dominant subgenome C whereas minor ones came from F followed by E and N. Their blue haplotype C+F matched the dominant subgenome C. Final results indicated that ‘*R. platycalpoides*’ (H1) is composed of subgenomes C and N, ‘*R. elatior*’ (H2) of C and F, ‘*R. pseudocassubicus*’ (H3) of C and E, and ‘*R. hungaricus*’ (H4) of C and F.

Moreover, we inferred varying subgenome contributions for the polyploid ‘*R. pilisiensis*’ (H7) of cluster II (RAD-Seq)/III (nuclear genes), but consensus supported by CP results revealed subgenome F as the dominant one. The likeliest scenario is an autopolyploid origin. The polyploid ‘*R. indecorus*’ (H8), positioned in cluster II (RAD-Seq)/IV (nuclear genes), showed three subgenomes, whereas N was the slightly dominant one, and C and E the minor ones. ‘*R. fissifolius*’ (H9) was characterized by the orange haplotype N and is also composed of three different subgenomes (E, F, and N).
Fig. 7 (a–h, left) Reconstructed phylogenetic networks based on restriction site-associated DNA sequencing (RAD-Seq), nuclear target enrichment genes (TEGs), and plastid data (CP). Final networks of allopolyploids are based on genetic structure and phylogenetic network results (consensus results) corrected by the full likelihood approach + Akaike information criterion (AIC) calculations in PHYLONET and CP data. P1 defines the largest subgenome contribution, followed by P2 and P3. The network topology follows the published rooted phylogeny of *Ranunculus auricomus* (*without tetraploid Ranunculus marsicus*; Karbstein et al., 2020c). Curves indicate subgenome contributions (P1–P3). (a–h, right) Bar charts based on SNPloid results show single nucleotide polymorphism (SNP) origins in percentages (cat1&2 = SNPs identical to DIPLOID2 or DIPLOID1/reference, cat3/4 = derived SNPs, cat5 = homeo-SNPs). Concerning H5 and H8, we calculated two SNPloid analyses because three parents have contributed to their origin. Coloring of sexual progenitor subgenomes is according to Fig. 2. Subgenome C = *Ranunculus cassubicifolius*, F = *Ranunculus flabellifolius*, E = *Ranunculus envallirensis*, and N = *Ranunculus notabilis*. 

© 2022 The Authors
New Phytologist © 2022 New Phytologist Foundation.

New Phytologist (2022) 235: 2081–2098
www.newphytologist.com
The polyploids ‘R × glechomoides’ (H₃), ‘R × subglechomoides’ (H₈), and ‘R × leptomeris’ (H₁₀) exhibited subgenome E as the dominant contribution. In most genetic structure analyses, these polyploids were also situated close to E. CP analyses showed the green haplotype E for H₆ and H₁₀, but not for H₉. Final results indicated E and F subgenome contributions for H₆ and H₁₀. ‘R × elation’ is closer to subgenome C indicating C subgenome dominance. We also illustrate characteristic basal leaf types of taxa during anthesis (variation not covered, two types illustrated for R. flabellifolius due to the occurrence of undivided and divided types, frequently observed in collections). The hybrid scheme is based on results of Table 1. Currently used taxon names of allopolyploids are given, but we specified the polyploids as nothotaxa because of known hybrid origin (Hörandl, 2022).

**Fig. 8** Hybrid scheme of sexual progenitors and selected apomictic polyploid Ranunculus auricomus derivates (H₁–H₁₀). The diploid sexual progenitor species Ranunculus cassinicifolius (C), Ranunculus flabellifolius (F), Ranunculus notabilis (N), Ranunculus envalirensis (E), and a hypothetical unknown taxon (U) in different combinations gave rise to apomictic polyploid derivates (Figs 2, 7). Differently dashed lines to the left and right specify parental subgenome contributions of alloplopoloids. Subgenome dominance is shown by the relative position of the polyploid to the progenitors, for example, ‘R × elation’ is closer to subgenome C indicating C subgenome dominance. We also illustrate characteristic basal leaf types of taxa during anthesis (variation not covered, two types illustrated for R. flabellifolius due to the occurrence of undivided and divided types, frequently observed in collections). The hybrid scheme is based on results of Table 1. Currently used taxon names of allopolyploids are given, but we specified the polyploids as nothotaxa because of known hybrid origin (Hörandl, 2022).

**Origin of SNPs**

The SNIPLOID analyses based on RAD-Seq data supported allopolyploid origins with 3–33% homeo-SNPs of cat3. H₁, H₃, and H₄ showed relatively high percentages of homeo-SNPs (> 20%), whereas H₂, H₅–H₁₀ exhibited low amounts (< 15%). The majority of SNPs, however, indicated considerable post-origin evolution of allopolyploids (Fig. 7a–h; Table S8). Across datasets, SNIPLOID assessed 64–93% interspecific SNPs of cat1&2, and 3–5% derived SNPs of cat3/4. Interspecific SNPs of cat1&2 were lowest for polyploids H₁, H₃, and H₄ and highest for H₂, H₅–H₁₀. The polyploids ‘R × glechomoides’ (H₃), ‘R × subglechomoides’ (H₈), and ‘R × leptomeris’ (H₁₀) exhibited subgenome E as the dominant contribution. In most genetic structure analyses, these polyploids were also situated close to E. CP analyses showed the green haplotype E for H₆ and H₁₀, but not for H₉. Final results indicated E and F subgenome contributions for H₆ and H₁₀. ‘R × elation’ is closer to subgenome C indicating C subgenome dominance. We also illustrate characteristic basal leaf types of taxa during anthesis (variation not covered, two types illustrated for R. flabellifolius due to the occurrence of undivided and divided types, frequently observed in collections). The hybrid scheme is based on results of Table 1. Currently used taxon names of allopolyploids are given, but we specified the polyploids as nothotaxa because of known hybrid origin (Hörandl, 2022).

**Discussion**

Evolution of neopolyploids is an emerging and bioinformatically challenging field, with important consequences for understanding plant speciation and subsequent macroevolution (Soltis et al., 2015; Landis et al., 2018; Rothfels, 2021). The applied workflow disentangled polyploid relationships, parental contributions, geographical patterns, and genomic compositions in the less than 1.0 Ma R. auricomus complex. Our results confirmed that allopolyploidy is the dominant polyploid formation type (Sochor et al., 2015; Dauphin et al., 2018; Hörandl, 2018; Rothfels, 2021), but also demonstrated substantial post-origin evolution, which is unique compared to other studies. The R. auricomus model system is thus the first well-studied large apomictic polyploid species complex using RRS genomic data. Several genomic datasets (RAD-Seq, nuclear genes, plastid data) along with a priori (multi-approach) information were combined with up-to-date bioinformatic tools into a comprehensive workflow – starting with diploid progenitors and ending with the origin of polyploid derivatives, to receive a complete picture of the underlying reticulate evolutionary processes.
Integration of datasets and analyses

Our study demonstrates that the combination of different RRS genomic datasets is necessary to resolve young reticulate relationships. RAD-Seq provided the highest number of genetic information, which is important to tackle among and within-species relationships of groups characterized by low genetic divergence, reticulations, and ILS. RADPAINTER tolerates moderate to high amounts of missing data and compares all SNPs independent of ploidy levels across loci and individuals (Malinsky et al., 2018; Wagner et al., 2021), which is important for allopolyploid analysis. Moreover, the employed SNMF algorithm easily takes mixed-ploidy datasets; it is not only faster, but also less sensitive to apomictic population structure and missing data than the popular STRUCTURE software (Frichot et al., 2014; Stift et al., 2019; Frichot & François, 2020). RADPAINTER and SNMF impressively showed hybridity of polyploids. However, drawbacks are the extraction of only the average evolutionary signal (RADPAINTER) or incorporation of less genetic information (SNMF). Therefore, analyses based on phased nuclear genes were conducted to allow more accurate inferences of reticulate polyploid relationships (Dauphin et al., 2018; Eriksson et al., 2018; Lautenschlager et al., 2020; Rothfels, 2021). Coalescent-based STACEY species delimitation compared to RADPAINTER more clearly delimited the genetic structure of the polyploid complex (Fig. 4a,b). STACEY takes allelic information into account, and was thus also able to find shared high posterior probability between allopolyploids and their sexual progenitor species. Compared to RADPAINTER and STACEY, limitations of phylogenetic networks (PhyloNetworks, PhyloNet) are the restriction to only two progenitors in hybrid modeling. Nevertheless, these analyses with subsequent allopolyploid vs autopolyploid origin testing based on phased nuclear genes (see also Tiley et al., 2021) were the most important parts to unravel final subgenome contributions of polyploids. Recently, methods have been developed able to assign alleles to subgenomes and infer advanced allopolyploid phylogenetic networks (Jones, 2017a; Freyman et al., 2020; Lautenschlager et al., 2020; Slender et al., 2021). These methods need subgenomes that are genetically well-differentiated, knowledge about diploid parental contributions, and/or small- to medium-sized datasets (<100 loci), which is still unfeasible for analyzing young non-model polyploid groups with large-scale genomic information.

RAD-Seq and Hyb-Seq datasets represent the genome and can be much more easily gained and analyzed at lower costs for a large number of samples than entire transcriptomes or genomes (McKain et al., 2018; Johnson et al., 2019). Comprehensive RAD-Seq and Hyb-Seq results were supported by well-resolved haplotype groups. CP data unraveled nuclear-plastid discordances and pinpointed maternal progenitors, which in most cases corresponded to the dominant progenitor subgenome. Performing several network methods across different datasets informed by plastid information (i.e. consensus making) is therefore the best way to get a reliable picture of young polyploid evolution. Moreover, dissecting genetic information (SNPs) of polyploids for post-origin processes using SNiPLOID provides crucial information about divergence and stability of lineages. A limitation of SNiPLOID is that the parental species must be defined for the input, only single samples can be analyzed, and that the algorithm is so far limited to tetraploids. However, results for the tested polyploids here were strikingly similar, suggesting that they reflect well the post-origin processes of this neopolyploid complex.

Genetic structure, nuclear-plastid discordance, and origin of polyploids

Despite predominant apomictic reproduction, young allopolyploid lineages interacted many times with their sexual progenitor species and each other, resulting in large, network-like relationships. Genetic structure analyses based on RAD-Seq and nuclear genes surprisingly revealed that all 83 included R. auricomus taxa were grouped in only 3–5 main clusters (Figs 4, 5). No clear cluster-specific morphological trend is recognizable, except that taxa with undivided basal leaves were predominantly found in cluster I. Each main cluster contained at least one sexual species surrounded by polyploid apomictic taxa, indicating that polyploids received a main genetic contribution from their cluster-specific diploid progenitor. This pattern is rarely found in other polyploid complexes, where usually groups with several diploid and (allo)polyploid taxa, but also groups with polyploids only were detected (Kirschner et al., 2015; Carter et al., 2019; Wagner et al., 2020). The observed main clusters are west–east distributed in Europe (Fig. 5a), each ranging from southern to northern Europe, respectively. This pattern probably reflects the allopatric distributions of sexual progenitors in combination with migration of populations due to past climatic changes (Abbott et al., 2013; Tomasello et al., 2020).

Moreover, the reticulate evolution of the R. auricomus complex is supported by several nuclear-plastid discordances. For example, the central European subcluster (Fig. 5b) widely corresponds to the gray ‘polyploid-only’ haplotype of the ML plastid tree and splitsgraph (Fig. 6a,b). Therefore, these polyploids possess a plastid type from an unknown diploid, suggesting an already extinct or unsampled maternal sexual progenitor. The fact that R. auricomus populations have been extensively studied for more than 80 yr (Koch, 1939; Borchers-Kolb, 1985; Hörandl & Gutermann, 1998) and ploidy levels have been well documented within the last decades in central Europe (Jalas & Suominen, 1989; Dunkel et al., 2018; Paule et al., 2018; Karbstein, 2021; Karbstein et al., 2021a) makes it unlikely that an extant diploid was simply overlooked. Extinction of sexuals is a commonly considered or observed phenomenon in young polyploid complexes shaped by past climatic fluctuations (Sochor et al., 2015; Rothfels, 2021) and is supported here by missing speciation events between 0.6 and 0.3 Ma (Tomasello et al., 2020).

The tested allopolyploids are composed of various subgenome combinations. The presence of multiple different copies in (allo) polyploids probably provides larger physiological and phenotypic flexibility to respond to different environmental conditions (Hörandl, 2006; Blaine Marchant et al., 2016; van de Peer et al., 2017; Karbstein et al., 2021a). This partly explains the larger geographic distribution of polyploid apomicts compared to their diploid progenitors (Bierzchudek, 1985; Hörandl, 2006; Burgess...
et al., 2014; Karbstein et al., 2021a). Our results revealed that the majority of polyploids were composed of two, some of three, subgenome contributions (Figs 3–7; Table 1). We detected only one polyploid without evidence of a reticulate evolutionary history: ‘R. × piliiensis’ (H7). This lineage might also represent a segmental allopolyploid, as autopolyploidy and allopolyploidy are connected by transitions (Comai, 2005; Spoelhof et al., 2017).

All diploid sexual progenitor species were involved in allopolyploid formation (Fig. 8). Polyploids were probably formed multiple times out of different progenitor combinations. Extant sexuals have restricted ranges and are separated by thousands of kilometers across Europe (Fig. 2). Nevertheless, all main genetic clusters are present in central Europe (Fig. 5a,b). Sexual progenitors might have repeatedly hybridized in this region during past interglacial times, resulting in multiple allopolyploidization ‘waves’ with varying subgenome contributions. The phenotypic diversity of polyploid R. auricomus biotypes with more than 840 described morphospecies is therefore probably formed from only four extant and at least one likely extinct diploid sexual progenitor species. Other studies already demonstrated that few diploid progenitors were capable of producing a magnitude of allopolyploids, for example in Botrychium (Dauphin et al., 2018), Rubus (Sochor et al., 2015; Carter et al., 2019), or Taraxacum (Kirschner et al., 2015).

Post-origin genome evolution

We observed in the proportions of SNPs/alleles a dominance of one subgenome over the other which has the consequence that allopolyploids are grouped close to one of their diploid progenitors (Figs 3–7; Table 1). Per main genetic cluster, allopolyploids were usually composed of a dominant intra-cluster subgenome and 1(−2) minor, varying inter-cluster subgenome(s), although trigenomic polyploids showed rather similar contributions. Plastid data supported the dominant subgenome contribution and, in some cases, unraveled an additional or extinct progenitor (Table 1). Subgenome expression and sequence dominance is a common feature of allopolyploid post-origin evolution (Blischak et al., 2018; Alger & Edger, 2020; Mason & Wendel, 2020). In our neopolyploids, we detected considerable proportions of interspecific, post-origin SNPs (62–93%), and only a minority of SNPs from hybridogenic origins (3–36%; Fig. 7). Subgenome dominance is probably caused by HEs, segregation after hybridization, and gene flow due to facultative sexuality of apomicts after polyploidization. Homoeologous exchange is able to transfer chromosome segments of one parent to the homoeolog of the other during crossovers at meiosis, resulting in genomically homoyzogous regions (Mason & Wendel, 2020). This mechanism could have slightly increased the proportions of cat1&2 interspecific SNPs in R. auricomus polyploids directly after their origin. In addition, since apomixis establishes only stepwise, the first diploid hybrid generations still exhibit predominant sexual reproduction (Barke et al., 2018, 2020), allowing for extensive Mendelian segregation and a substantial increase of cat1&2 interspecific SNPs. However, after shift to obligate apomixis, HE and Mendelian segregation are no longer effective and hence explain probably not all of the observed substantial subgenome dominance. Facultative sexuality and maintenance of functional pollen probably allowed backcrossing between the newly formed polyploid hybrids and their parental species, and among polyploids. Varying degrees of observed facultative sexuality under natural conditions (mean 2%, range 0–34%; Karbstein et al., 2021a) potentially enabled these scenarios. In apomictic Rubus or Pilosella polyploids, which are also characterized by highly variable, facultative sexuality, post-origin scenarios of interlineage gene flow and backcrossing to sexual parents have also been considered (Sochor et al., 2015; Hörandl, 2018; Nardi et al., 2018; Carter et al., 2019).

Within main clusters, the majority of described polyploid morphospecies are intermingled and nonmonophyletic (except for H2, H6, H7; Figs S11–S15). Support metrics of these trees showed highly conflicting signals at middle and terminal branches, suggesting that cladogenetic (bifurcating) speciation from a polyploid ancestor is unlikely due to the presence of many reticulation processes. According to Grant’s (1981) definition, the R. auricomus complex is in an early mature stage of evolution, with extant diploid progenitors and a broad array of apomictic hybrid biotypes, but still without stable lineages. This hypothesis is confirmed by the low proportion of lineage-specific SNPs (2–5%; Fig. 7), which is very low compared to similarly aged or older allopolyploids (Gordon et al., 2020; Wagner et al., 2020). Genetic information on lineage characteristics and stability is crucial for the classification and delimitation of species (Grant, 1981; Hörandl, 2018), and gained knowledge of this study can be used in future comprehensive taxonomic revisions of the large R. auricomus species complex.

Acknowledgements

The authors acknowledge Franz G. Dunkel for providing garden plants and herbarium specimens, Ena Lehtsaar and Julius Schmidt for technical help, and John Paul Bradican for suggestions on previous manuscript versions. The authors thank the herbaria of Jena (JE), Munich (M), Oslo (O), Uppsala (UPS), and the University of Vienna (WU) for loans of R. auricomus type species material. The authors thank the referees and the editors for valuable comments on the manuscript. The work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft DFG, grant no. Ho4395/10-1) to EH within the priority program ‘Taxon-Omics: New Approaches for Discovering and Naming Biodiversity’ (SPP 1991).

Author contributions

KK, ST, and EH designed research; KK, LH, and EH collected plant materials; KK, ST, PM, BHB, and CP performed laboratory work; KK, ST, and NW analyzed data; KK wrote the article with contributions from all authors.

ORCID

Birthe Hilkka Barke https://orcid.org/0000-0001-5379-6055
Elvira Hörandl https://orcid.org/0000-0002-7600-1128
Ladislav Hodáč https://orcid.org/0000-0002-6885-1317
Data availability

The authors declare that basic data supporting the findings are available within the manuscript and Supporting Information. RAD-Seq, target enrichment, and CP alignments, and tables and figures supporting the results are deposited on FigShare (https://doi.org/10.6084/m9.figshare.14046305). RAD-Seq and target enrichment reads are deposited on the National Center for Biotechnology Information Sequence Read Archive (SRA): BioProject ID PRJNA627796, https://www.ncbi.nlm.nih.gov/bioproject/627796 and BioProject ID PRJNA628081, https://www.ncbi.nlm.nih.gov/bioproject/628081, respectively. Flow cytometric (FC) and flow cytometric seed screening (FCSS) data are also stored in FigShare (https://doi.org/10.6084/m9.figshare.13352429). The authors deposited custom bash, R, and Julia scripts on Github (https://github.com/KK260/Ranunculus_auricomus_phylogenetic_network_scripts).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Branch JD, Medrano JF, et al. 2018. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: 1–7.

Barke BH, Daubert M, Hoda H. 2014. Effects of apomixis and polyploidy on diversification and geographic distribution in Amelanchier (Rosaceae). American Journal of Botany 101: 1375–1387.

Barke BH, Liston A, Bassil NV, Alice LA, Bushakra JM, Sutherland BL, Mockler TC, Bryant DW, Hummer KE. 2019. Target capture sequencing unravels Rubus evolution. Frontiers in Plant Science 10: 1–18.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6: 836–846.

Coyne J, Orr H. 2004. Speciation. Sunderland, MA, USA: Sinauer.

Dann M, Bellot S, Schepella S, Schafer H, Tellier A. 2017. Mutation rates in seeds and seed-banking influence substitution rates across the angiosperm phylogeny. bioRxiv. doi: 10.1101/156398.

Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTEST-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Molecular Biology and Evolution 37: 291–294.

Dauphin B, Grant JR, Farrant DR, Rothfels CJ. 2018. Rapid allopolyploid radiation of moonwort ferns (Barychodium, Ophioglossaceae) revealed by PacBio sequencing of homologous and homologous nuclear regions. Molecular Phylogenetics and Evolution 120: 342–353.

Davey JW, Hohenlohe PA, Etter PD, Boone IQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12: 499–510.

Dunkel FG, Gregor T, Paule J. 2018. New diploid species in the Ranunculus auricomus complex (Ranunculaceae) from W and SE Europe. Willdenowia 48: 227–257.

Eaton DAR, Overcast I. 2020. IPRAD: interactive assembly and analysis of RAD-Seq datasets. Bioinformatics 36: 2592–2594.

Eaton DAR, Spprigs EL, Park B, Donoghue MJ. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66: 399–412.

Eriksson JS, de Sousa F, Bertrand YJK, Antonelli A, Oxloman B, Pfeld BE. 2018. Allele phased is critical to revealing a shared allopolyploid origin of Medicago and M. strasius (Fabaceae). Molecular Evolutionary Biology 18: 9.

Feher J, Slavikova R, Pastova L, Josefova J, Mraz P, Chrtek J, Bertrand YJK. 2021. Molecular evolution and organization of ribosomal DNA in the hawkweed tribe Hieracinae (Cichorieae, Asteraeae). Frontiers in Plant Science 12: 23.

Fer T, Schmickl RE. 2018. HyPhyMager: target enrichment data analysis from raw reads to species trees. Evolutionary Bioinformatics 14: 1–9.

Folk RA, Mandel JR, Freudenstein JV. 2015. A protocol for targeted enrichment of intron-containing sequence markers for recent radiations: a phylogenetic example from Heuchera (Saxifragaceae). Applications in Plant Science 3: 1500039.

Freeland JR, Kirk H, Petersen SD. 2011. Molecular ecology. Chichester, UK: Wiley.

Freyman WA, Johnson MG, Rothfels CJ. 2020. HOMOLOGIZER: phylogenetic phasing of gene copies into allopolyploid subgenomes. bioRxiv. doi: 10.1101/2020.10.22.351486.

Frichot E, Francois O. 2015. LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution 6: 925–929.

Frichot E, Francois O. 2020. R package for landscape and ecological association studies. v.3.0.0. [WWW document] URL https://bioconductor.org/packages/release/bioc/html/LEA.html [accessed 1 December 2020].

Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O. 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196: 973–983.

Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Faircloth BC, Antonelli A. 2019. Phylogenetic enrichment reads are deposited on the National Center for Biotechnology Information Sequence Read Archive (SRA): BioProject ID PRJNA627796, https://www.ncbi.nlm.nih.gov/bioproject/627796 and BioProject ID PRJNA628081, https://www.ncbi.nlm.nih.gov/bioproject/628081, respectively. Flow cytometric (FC) and flow cytometric seed screening (FCSS) data are also stored in FigShare (https://doi.org/10.6084/m9.figshare.13352429). The authors deposited custom bash, R, and Julia scripts on Github (https://github.com/KK260/Ranunculus_auricomus_phylogenetic_network_scripts).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R et al. 2013. Hybridization and speciation. Journal of Evolutionary Biology 26: 229–246.

Alger EL, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. Current Opinion in Plant Biology 54: 108–113.

Andermann T, Fernandes AM, Olsson U, Töpel M, Pfeld B, Oxloman B, Alexio A, Faircloth BC, Antonelli A. 2019. Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. Systematic Biology 68: 32–46.

Asker S, Jerling L. 1992. Apomixis in plants. Boca Raton, FL, USA: CRC Press.

Babcock EB, Stebbins GL. 1938. The American species of Crops. Their interrelationships and distribution as affected by polyploidy and apomixis. Washington, DC, USA: Carnegie Institution of Washington.

Barke BH, Daubert M, Hörandl E. 2018. Establishment of apomixis in diploid F$_2$ hybrids and inheritance of apomixy in the Ranunculus auricomus complex. Frontiers in Plant Science 9: 1–12.

Barke BH, Karstein K, Daubert M, Hörandl E. 2020. The relation of meiotic behaviour to hybridity, polyploidy and apomixis in the Ranunculus auricomus complex (Ranunculaceae). BMC Plant Biology 20: 523.

Bierzychudek P. 1985. Patterns in plant parthenogenesis. Experientia 41: 1255–1264.

Blaine Marchant DB, Solits DE, Solits PS. 2016. Patterns of abiotic niche shifts in alloploids relative to their progenitors. New Phytologist 212: 708–718.

Blüschak PD, Mabey ME, Conant GC, Pires JC. 2018. Integrating networks, phylogenomics, and population genomics for the study of polyploidy. Annual Review of Ecology, Evolution, and Systematics 49: 253–278.

Borchers-Kolb E. 1985. Ranunculus sect. Auricomus in Bayern und den angrenzenden Gebieten II. Spezieller Teil. Mitteilungen der Botanischen Staatsammlung München 21: 49–300.

Borowicz ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4: e1660.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drumond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537.
Hodac L, Klatt S, Hojsgaard D, Sharbel TF, Horandl E. 2019. A little bit of sex prevents mutation accumulation even in apomictic polyploid plants. BMC Evolutionary Biology 19: 170.

Hodac L, Scheben AP, Hojsgaard D, Paun O, Horandl E. 2014. ITS polymorphisms shed light on hybrid evolution in apomictic plants: a case study on the Ranunculus auricomus complex. PLoS ONE 9: 28–30.

Hojsgaard D, Horandl E. 2019. The rise of apomixis in natural plant populations. Frontiers in Plant Science 10: 358.

Horandl E. 2006. The complex causality of geographical parthenogenesis. New Phytologist 171: 525–538.

Horandl E. 2018. The classification of sexual organisms: old myths, new facts, and a novel pluralistic approach. Taxon 67: 1066–1081.

Horandl E. 2022. Novel approaches for species concepts and delimitation in polyploids and hybrids. Plants 11: 204.

Horandl E, Greilhuber J, Klimova K, Paun O, Temsch E, Emadzade K, Hodalova I. 2009. Reticulate evolution and taxonomic concepts in the Ranunculus auricomus complex (Ranunculaceae): insights from analysis of morphological, karyological and molecular data. Taxon 58: 1194–1215.

Horandl E, Gutermann W. 1998. Der Ranunculus auricomus – Komplex in Österreich 2. Die R. caucasicus we do not have information about the maternal progenitor of a bunch of polyploidy apomicts R. monophyllus- und R. fallax-Sammelgruppe. Botanische Jahrbücher für Systematik, Pflanzengeographie und Pflanzenkunde 120: 545–598.

Huang DL, Hefer CA, Kolosova N, Douglas CJ, Cronk QCB. 2014. Whole plastome sequencing reveals deep plastid divergence and cytonuclear discordance between closely related balsam poplars, Populus balsamifera and P. trichocarpa (Salicaceae). New Phytologist 204: 693–703.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254–267.

Jalas J, Suominen J. 1989. Atlas florae Europaeae. Distribution of vascular plants in Europe, vol. 8, Nymphaeaceae to Ranunculaceae. Helsinki, Finland: The Committee for Mapping the Flora of Europe, Societas Biologica Fennica, Vanamo.

Jaron KS, Bast J, Nowell RW, Ranallo-Benavidez TR, Robinson-Rechavi M, Schwander T. 2020. Genomic features of parthenogenetic animals. Journal of Heredity 112: 19–33.

Johnson MG, Pokorny L, Dodsworth S, Botigue LR, Cowan RS, Devault A, Karbstein K, Tomasello S, Hoda H, Prinz K. 2019. Desert-like badlands and surrounding (semi-)dry grasslands of Central Germany promote small-scale phenotypic and genetic differentiation in Thymus praecox. Ecology and Evolution 9: 14066–14084.

Kearse M, Moir R, Wilson A, Stones-Havas S, Churug M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. 2012. Generous Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Kirschner J, Záveska Drabklová L, Steptánek J, Uhlemann I. 2015. Towards a better understanding of the Taraxacum evolution (Compositae–Chicorieae) on the basis of mtDNA of sexually reproducing species. Plant Systematics and Evolution 301: 1135–1156.

Koch W. 1939. Zweiter Beitrag zur Kenntnis des Formenkreises von Ranunculus auricomus L.; Studien über kritische Schweizerpflanzen III. Berichte der Schweizerischen Botanischen Gesellschaft 49: 541–554.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35: 4453–4455.

Landis JB, Soltis DE, Zheng L, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. American Journal of Botany 105: 348–363.

Lautenschlager U, Wagner F, Oberprieler C. 2020. ALLCoPox: inferring allele co-ancestry in polyploids. BMC Bioinformatics 21: 1–9.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Z L, Melkonian M, Mirabab S et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574: 679–685.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.

Malinsky M, Trbuch E, Lawson DJ, Falush D. 2018. Radpainter and FinerRadstructure: population inference from RAD-seq data. Molecular Biology and Evolution 35: 1284–1290.

Mason AS, Wendel JF. 2020. Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Frontiers in Genetics 11: 1014.

McDade LA. 1992. Hybrids and phylogenetic systematics. 1. The impact of hybrids on cladistic analysis. Evolution 46: 1329–1346.

McDade LA. 1995. Hybridization and phylogenetics. In: Hoch PC, Stephenson A, eds. Experimental and molecular approaches to plant systematics. St Louis, MO, USA: Missouri Botanical Garden, 305–331.

McKain MR, Johnson MG, Urbe-Consvers S, Eaton D, Yang Y. 2018. Practical considerations for plant phylogenomics. Applications in Plant Sciences 6: 15.

Melichárková A, Senker M, Zozomová-Líhová J, Skokanová K, Singjarlová B, Kacimárová T, Cabonová M, Kempa M, Šrámková G, Mandáková T et al. 2020. So closely related and yet so different: strong contrasts between the evolutionary histories of species of the Cardamine pratensis polyplody complex in Central Europe. Frontiers in Plant Science 11: 1988.

Nardi FD, Dobes C, Muller D, Grasegger T, Myllynen T, Alonso-Moros H, Trisch S. 2018. Sexual intraspecific recombination but not de novo origin governs the genesis of new apomictic genotypes in Potentilla puberula (Rosaceae). Taxon 67: 1108–1131.

Olave M, Meyer A. 2020. Implementing large genomic single nucleotide polymorphism data sets in phylogenetic network reconstructions: a case study of particularly rapid radiations of Cichlid fish. Systematic Biology 69: 848–862.

Otto SP, Whitten J. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34: 401–437.

Paule J, Dunkel FG, Schmidt M, Gregor T. 2018. Unraveling phylogenetic relationships, reticulate evolution, and genome composition of polyploid plant complexes by RAD-Seq and Hyb-Seq. bioRxiv: doi: 10.1101/2021.08.30.458250.

Karstlein K. 2019. The complex causality of geographical parthenogenesis. New Phytologist 171: 223–236.

Karstlein K, Tomasello S, Hodac L, Wagner N, Marinéck P, Barke BH, Pätzold C, Horandl E. 2021b. Moving beyond assumptions: polyploidy and environmental effects explain a geographical parthenogenesis scenario in European plants. Molecular Ecology 30: 2659–2675.

Karstlein K, Tomasello S, Hodac L, Horandl E. 2021a. Polyploidy and environmental effects explain a geographical parthenogenesis scenario in European plants. Molecular Ecology 30: 2659–2675.
Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. American Journal of Botany 105: 385–403.

del de Peer Y, Mizrachi E, Marchal K, Soltis PS, Marchant DB, van de Peer Y, Soltis DE. 2015. A novel approach to resolving the origin of allotetraploids using Hyb-Seq data: a case study of the Balkan mountain endemic novel relationships among major lineages. American Journal of Botany 102: 1753–1756.

Welch DM, Meselson M. 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science 288: 1211–1215.

Wen D, Yu Y, Zhu J, Nahkleh L. 2018. Inferring phylogenetic networks using PhyloNet. Systematic Biology 67: 735–740.

Wendel JF. 2015. The wondrous cycles of polyploidy in plants. American Journal of Botany 102: 1753–1756.

Wicke M, Weigand H, Weigand AM, Leese F. 2018. Genome- wide single- nucleotide polymorphism data reveal cryptic species within cryptic freshwater snail species – the case of the Ancylus fluviatilis species complex. Ecology and Evolution 8: 1063–1072.

Wittendorf K, Krause S, Hobaek T, Leese F, Schmid M, Schlicht M, McDonnell A. 2021. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. Applications in Plant Science 2: 1400042.

Zhang C, Rabiee M, Sayaeei E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19: 153.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Clustered fineRADSastructure coancestry matrices.

Fig. S2 Cross-entries based on SNMF results.

Fig. S3 Ancestry proportions based on SNMF analyses.

Fig. S4 Geographic maps showing ancestry coefficients across Europe (K = 3–5).

Fig. S5 Geographic maps showing separate ancestry coefficients across Europe (K = 3, min30, see Figs 5, S3b, S4b).

Fig. S6 Geographic maps showing separate ancestry coefficients across Europe (K = 4, min30, see Figs S3b, S4b).

Fig. S7 Geographic maps showing separate ancestry coefficients across Europe (K = 5, min30, see Figs S3b, S4b).

Fig. S8 Geographic maps showing genetic clusters and ancestry coefficients across Europe.

Fig. S9 Concatenated ML tree inferred from plastid regions (CP).

Fig. S10 Neighbor-net analysis (SplitsTree) of plastid (CP) data.
**Fig. S1** Phylogenetic trees based on RAD-Seq and target enrichment data.

**Fig. S12** Concatenated ML tree with summarized main clades (min30).

**Fig. S13** Concatenated ML tree with summarized main clades (min50).

**Fig. S14** Concatenated ML tree with quartet sampling scores (QS).

**Fig. S15** Coalescent-based tree ASTRAL tree.

**Notes S1** HYBPHYLOMAKER settings (target enrichment analysis).

**Notes S2** Loci selection (target enrichment analysis).

**Notes S3** Further sample filtering for target enrichment data analysis.

**Notes S4** Detailed STACEY settings (target enrichment data analysis).

**Table S1** Location details of sampled *Ranunculus auricomus* populations across Europe.

**Table S2** Quality trimming and read mapping (target enrichment analysis).

**Table S3** Loci selection (target enrichment analysis).

**Table S4** Results of the phasing procedure (target enrichment analysis).

**Table S5** Read mapping (plastome analysis).

**Table S6** List of regions used for the chloroplast phylogeny (plastome analysis).

**Table S7** Genetic structure and phylogenetic network results of tested tetraploid *Ranunculus auricomus* accessions (H1–H10).

**Table S8** SNP discovery based on RAD-Seq SNiPLOID results.

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.