Received: 2 August 2021 | Revised: 4 January 2023 | Accepted: 5 January 2023

DOI: 10.1111/mec.16844

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Genomic evidence for contrasting patterns of host-associated genetic differentiation across shared host-plant species in leaf- and bud-galling sawflies

Craig T. Michell[1] | Natascha Wagner[2] | Marko Mutanen[3] | Kyung Min Lee[3] | Tommi Nyman[4]

[1]Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu, Finland

[2]Department of Systematics, Biodiversity and Evolution of Plants (with Herbarium), University of Goettingen, Göttingen, Germany

[3]Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland

[4]Department of Ecosystems in the Barents Region, Norwegian Institute of Bioeconomy Research, Svanvik, Norway

Correspondence
Craig T. Michell, Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), 23955, Thuwal, Saudi Arabia.
Email: craig.michell@kaust.edu.sa

Present address
Kyung Min Lee, Zoology Unit, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

Funding information
Academy of Finland, Grant/Award Number: 294466

Handling Editor: Tatiana Giraud

Abstract

Resource specialization and ecological speciation arising through host-associated genetic differentiation (HAD) are frequently invoked as an explanation for the high diversity of plant-feeding insects and other organisms with a parasitic lifestyle. While genetic studies have demonstrated numerous examples of HAD in insect herbivores, the rarity of comparative studies means that we still lack an understanding of how deterministic HAD is, and whether patterns of host shifts can be predicted over evolutionary timescales. We applied genome-wide single nucleotide polymorphism and mitochondrial DNA sequence data obtained through genome resequencing to define species limits and to compare host-plant use in population samples of leaf- and bud-galling sawflies (Hymenoptera: Tenthredinidae: Nematinae) collected from seven shared willow (Salicaceae: *Salix*) host species. To infer the repeatability of long-term cophylogenetic patterns, we also contrasted the phylogenies of the two galler groups with each other as well as with the phylogeny of their *Salix* hosts estimated based on RADseq data. We found clear evidence for host specialization and HAD in both of the focal galler groups, but also that leaf gallers are more specialized to single host species compared with most bud gallers. In contrast to bud gallers, leaf gallers also exhibited statistically significant cophylogenetic signal with their *Salix* hosts. The observed discordant patterns of resource specialization and host shifts in two related galler groups that have radiated in parallel across a shared resource base indicate a lack of evolutionary repeatability in the focal system, and suggest that short- and long-term host use and ecological diversification in plant-feeding insects are dominated by stochasticity and/or lineage-specific effects.

KEYWORDS
co-evolution, host-associated genetic differentiation, insect–plant interactions, parasites, replicated evolution, speciation

1 | INTRODUCTION

Nearly half of all animal species on Earth are parasites (Weinstein & Kuris, 2016; Windsor, 1998). The parasitic lifestyle is defined by how consumer individuals utilize resources consisting of other (host) organisms: each parasite individual feeds exclusively on or within a single host individual during at least one life stage (Lafferty et al., 2015). Species-rich and ecologically central groups falling under this definition are, for example, ecto- and endoparasites of animals, but also most insects that feed on plants (Kawecki, 1998; Nylin et al., 2018). The intimate relationship between parasites and hosts fundamentally influences ecological phenomena such as food-web and population dynamics (Lafferty et al., 2015), microevolutionary forces acting on niche specialization and resource–consumer co-evolution (Kawecki, 1998), as well as macro-evolutionary trajectories, including trait evolution and the balance between speciation and extinction rates (Brooks & McLennan, 2002; Hay et al., 2020; Hembry & Weber, 2020; Mayhew, 2018; Weber et al., 2017).

As shown by Weinstein and Kuris (2016), the high species richness of parasitic taxa appears to reflect a combination of frequent adoption of the parasitic mode of life and extraordinarily rapid rates of net diversification within some groups. An oft-invoked explanation for elevated diversification rates in plant-feeding insects and other parasitic groups is ecological speciation through host-associated genetic differentiation (HAD)—the emergence of genetically divergent lineages as a result of adaptation to alternative host species and taxa (Boyd et al., 2022; Drès & Mallet, 2002; Forbes et al., 2017; Nosil, 2012; Nylin et al., 2018; Schluter, 2000). While extant parasitic species tend to be distinctly specialized in their use of available hosts, frequent temporal and topological discordances in the phylogenies of hosts and parasites (Bell et al., 2018; Jousselin et al., 2013; Leppänen et al., 2012; Lopez-Vaamonde et al., 2006; Percy et al., 2004; Scheffer et al., 2021) show that parasites occasionally shift their hosts through periodic expansions and restrictions in niche use (Janz & Nylin, 2008; Nylin et al., 2018). HAD constitutes the second step in ecological speciation driven by such expansion–restriction cycles, and is thought to arise as a result of contrasting selective pressures acting on traits that influence finding of, or performance on, hosts that differ in their ecological, morphological, physiological or chemical defensive traits (Itami & Craig, 2008; Nosil, 2012). Genetic surveys have revealed frequent HAD in animal parasites (Bell et al., 2018; Galbreath & Hoberg, 2015; Simmonds et al., 2020) and many groups of insect herbivores (Forbes et al., 2017). In the latter, classic examples include apple maggot flies (Hood et al., 2020), yucca moths (Drummond et al., 2010), pea aphids (Peccoud et al., 2009) and goldenrod ball-gall flies (Stireman 3rd et al., 2005).

While specialized host use and HAD are frequent in the animal kingdom, the wide specialist–generalist continuum often observed even among related species (e.g., Hardy et al., 2020; Kuchta et al., 2020; Martinů et al., 2015; Nakadai & Kawakita, 2016; Scheffer et al., 2021) raises the question of how deterministic ecological speciation driven by HAD is (Althoff, 2008; Forbes et al., 2017; Itami & Craig, 2008; Johnson et al., 2012; Stireman 3rd et al., 2005). The question of evolutionary repeatability can also be extended to whether the dynamics of host use and diversification can be predicted over longer, phylogenetic timescales (Bell et al., 2018; Braga et al., 2020; Hamerlinck et al., 2016; Sweet et al., 2016). In the simplest scenario, concordant phylogenetic patterns across co-occurring but independently radiating parasite lineages could result from parallel cospeciation with their shared hosts. However, phylogenetic concordance across parasite radiations could also arise if the colonization of new hosts is favoured by chemical, ecological or morphological similarity to the current hosts (Becerra & Venable, 1999), or if geographical proximity of alternative hosts promotes colonization (Boyd et al., 2022; Calatayud et al., 2016; Percy et al., 2004). In the presence of biased colonization probabilities, successive colonization–speciation cycles could lead to concordant phylogenetic patterns across co-occurring parasite lineages even in cases in which concordance with the hosts would be absent (cf. Cruaud & Rasplus, 2016; de Vienne et al., 2007; Sweet et al., 2016).

In this respect, the high taxonomic and ecological diversity of herbivorous insects provide ample opportunities for studying the relative importance of stochasticity vs. determinism in HAD, speciation and niche-use dynamics: most plant groups host numerous insect lineages that have colonized them independently, and such parallel radiations essentially constitute evolutionary replicates of the same colonization and diversification process. Unfortunately, comparative analyses of HAD are still rare and have been constrained by a lack of efficient genetic marker systems, as such comparisons necessarily must target groups containing multiple very closely related taxa (Dorchin et al., 2015; Forbes et al., 2017; Mlynarek & Heard, 2018). This obstacle has only recently been removed by rapid developments in high-throughput sequencing methods and computation-intensive bioinformatics, which in combination provide unprecedented possibilities for studying HAD in nonmodel organisms for which genomic resources have hitherto not been available (Cerca et al., 2021; Johnson, 2019; Poveda-Martínez et al., 2020, 2022; Wachi et al., 2018).

Here, we applied population-genomic and phylogenomic approaches to test for similarities in the levels of HAD as well as long-term patterns of host use across seven shared willow (*Salix*) host species in leaf- and bud-galling sawflies belonging to the nematine sawfly subtribe Euurina (Hymenoptera: Tenthredinidae: Nematinae; Figure 1). In an eco-evolutionary sense, galling sawflies are archetypal parasites, because their larvae are confined to life inside galls induced on willow tissues by plant hormone analogues produced by ovipositing females and the larvae themselves (Yamaguchi et al., 2012). Both leaf and bud gallers are common, and both types of gall are found on numerous *Salix* species growing in the Holarctic region (Kopelke, 1999; Liston et al., 2017). However, while it is generally recognized that both galler groups consist of multiple species or host races, inferring the number of distinct lineages and their host-use patterns has proven difficult: Kopelke (1999) generally postulated a strict 1:1 relationship between hosts and gallers, so that each *Salix* species is utilized by a single galler species of each gall type. By contrast, subsequent studies have

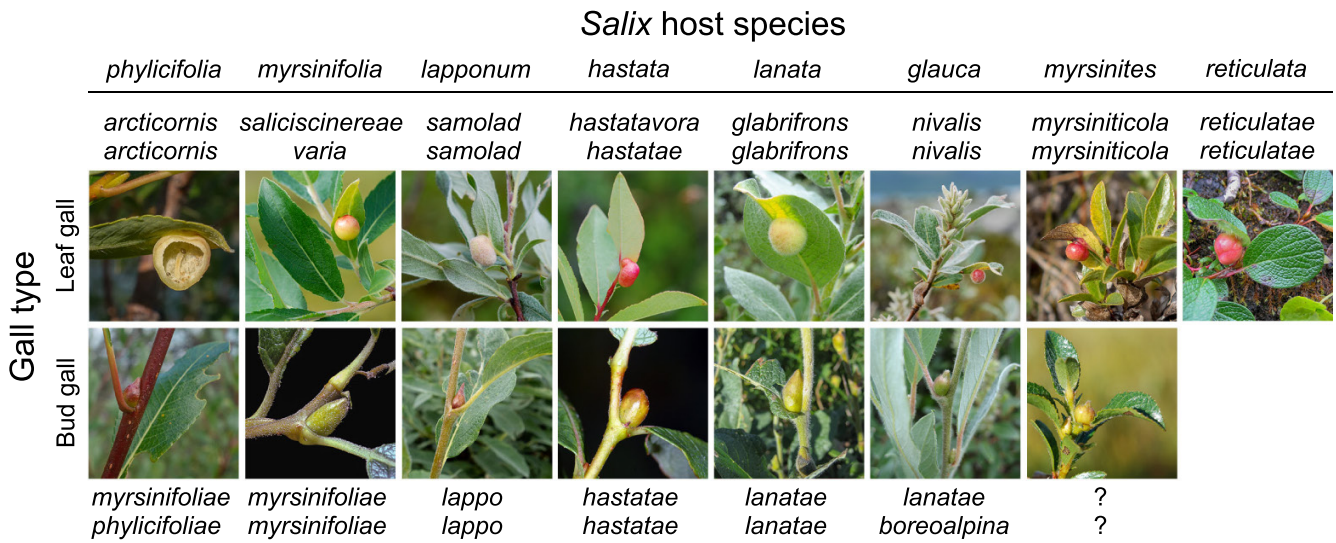


FIGURE 1 Photographs of galls induced by willow-galling sawflies on the eight focal *Salix* host species. The top row of images shows leaf galls; the gall on *S. phylicifolia* has been opened to show the sawfly larva inside. The bottom row shows various bud galls on the corresponding host species (note that bud gallers are not present on the creeping tundra willow *S. reticulata*). The two rows of species names above leaf galls and below bud galls denote species limits proposed by Liston et al. (2017; top rows) and Kopelke (1999; bottom rows; see Table S1). Photographs of bud galls on *S. myrsinifolia* and *S. hastata* by J.-P. Kopelke, others by T. Nyman

found that many of Kopelke's presumed specialist lineages cannot be separated based on morphological traits (Liston et al., 2017; Vikberg & Zinovjev, 2006), and genetic studies (Leppänen et al., 2014) have likewise suggested that at least some galler species are able to utilize multiple *Salix* hosts (Figure 1). In the setup of our study, we took advantage of the fact that the focal *Salix* species are hosts to both leaf- and bud-galling sawfly species. We first evaluated the existence of convergent patterns in species-level host use and HAD in the two galler clades. For this, we reared population samples of leaf- and bud-galling sawflies from the focal *Salix* hosts, and then used genome-level nuclear single nucleotide polymorphism (SNP) markers and mitochondrial DNA (mtDNA) sequences obtained through whole-genome resequencing to delimit galler species and to contrast the levels of host specificity and HAD across the two galler groups. Next, we expanded our focus to parallelisms in host-shift patterns on phylogenetic timescales. For this, we first constructed a phylogenetic tree for the *Salix* hosts based on restriction site-associated DNA sequencing (RADseq) data (Wagner et al., 2018, 2020) and tested for congruence between the species-level phylogenies of the gallers and their *Salix* hosts. Finally, because evolutionary repeatability in long-term patterns of host use can also arise through shared biases in host colonization, we also contrasted the phylogenies of the two galler groups directly with each other.

2 | MATERIAL AND METHODS

2.1 | Study system

Willows (*Salix* spp.) and sawflies in the subtribe Euurina (Hymenoptera: Tenthredinidae: Nematinae) that induce galls on willows constitute a species-rich and ecologically diverse model system for studying a multitude of eco-evolutionary questions. From a research perspective, a main practical benefit is that both willows and willow gallers are ubiquitous in most boreal, subarctic and arctic habitats across the Northern Hemisphere. Questions targeted so far range from the evolution of gall morphology and host use (Nyman et al., 2000) to the ecology of multitrophic networks (Gravel et al., 2019; Nyman et al., 2007, 2015; Price & Clancy, 1986), mechanisms of gall induction (Yamaguchi et al., 2012), and structure and function of larval microbiomes (Michell & Nyman, 2021).

At 400–450 species, willows form one of the most species-rich and ecologically important plant groups of the Holarctic region (Argus, 1997; Skvortsov, 1999). *Salix* species are utilized by a rich community of insect and mammalian herbivores, which runs contrary to the fact that willows produce a wide variety of defensive compounds, mainly phenolic glycosides, that vary in concentration and composition both within and across species (Julkunen-Tiitto, 1989; Nyman & Julkunen-Tiitto, 2005; Volf et al., 2015). A considerable obstacle for co-evolutionary research on *Salix*–herbivore systems has been the complex taxonomy of the genus. However, where attempts based on morphological traits and traditional genetic markers for decades failed to produce a stable phylogeny for *Salix* (e.g., Leskinen & Alström-Rapaport, 1999; Percy et al., 2014; Wu et al., 2015), Wagner et al. (2018, 2020) have recently been able to resolve interspecific relationships by using RADseq data.

Willow-galling sawflies form a monophylum of more than 200 species that induce either leaf folds or closed galls on salicaceous plants, but the species inducing closed galls are exclusively restricted to *Salix* species (Kopelke, 1999; Liston et al., 2017). Galling sawflies have traditionally been split into the genera *Phyllocolpa* (leaf folders and rollers), *Pontania* and/or *Eupontania* (leaf gallers), and *Euura* (petiole, shoot and bud gallers; e.g., Kopelke, 1999;

Vikberg & Zinovjev, 2006). However, some recent papers (e.g., Liston et al., 2017) have applied the generic classification of Prous et al. (2014), in which all sawfly gallers are included within a broad definition of the genus *Euura* (hereafter *Euura s.l.*), which also encompasses many nongalling groups of nematine sawflies (Table S1).

Molecular-phylogenetic analyses have demonstrated that the current diversity of willow-galling sawflies is a result of multiple temporally overlapping radiations of species inducing different galls (Liston et al., 2017; Nyman et al., 2000), so that many *Salix* species host multiple galler species. However, although sequence-based studies have provided a clear view of the overall phylogeny of the subtribe Euurina, inferring interrelationships and even the number of species within closely related clades have proven very difficult. The difficulties mainly stem from morphological uniformity within species-groups (Kopelke, 1999; Liston et al., 2017; Vikberg & Zinovjev, 2006) and extensive barcode and allele sharing between closely related species (Leppänen et al., 2014; Liston et al., 2017; Nyman, 2002). However, whether the complex genetic patterns found with traditional marker systems are a result of erroneous morphology- and host-based species definitions, incomplete lineage sorting, or hybridization across porous species boundaries remains unknown.

2.2 | Sample collection

2.2.1 | Gallers

Hymenopterans employ a haplodiploid sex determination system, with males being haploid and females diploid (Blackmon et al., 2017). Sawfly larvae cannot be reliably sexed, so in order to obtain diploid females for our population-genomic analyses, we collected leaf and bud galls from *Salix* hosts and reared the larvae to adults following protocols in Nyman (2002) and Nyman et al. (2015) for bud and leaf gallers, respectively. In brief, late-stage galls containing sawfly larvae were collected from identified *Salix* hosts in the field, while maintaining a general spacing of at least 20 m between galls collected from a particular host species. This sampling scheme minimizes the likelihood of sampling related galler individuals and ensures that galls from different host species are generally intermixed with respect to space. After overwintering, emerging adults were sexed and frozen individually in Eppendorf tubes at −80°C and later −20°C (bud gallers) or placed individually in 99.5% ethanol in 2-ml screw-cap tubes and stored at −20°C (leaf gallers). Bud galls were collected from seven willow species (*Salix phylicifolia*, *S. myrsinifolia*, *S. lapponum*, *S. hastata*, *S. lanata*, *S. glauca* and *S. myrsinites*) in the vicinity of Kilpisjärvi, Finland (69°02'34.8"N, 20°48'07.2"E; Table S1). Leaf gallers were collected from the same willow hosts from the vicinity of Abisko in Sweden (68°21'18.0"N, 18°48'57.6"E), but with two differences in relation to bud gallers: first, leaf galls were collected also from the creeping tundra willow *S. reticulata*, which is not used by bud gallers. Second, because leaf galls on *S. myrsinites* were rare in the sampling year, only three females were reared to adults, while nine larvae from this host were placed directly in alcohol. The sex of the latter individuals could therefore be determined only after sequencing but, based on ploidy inferred from their levels of heterozygosity, all of these specimens were also females (see below). Kilpisjärvi and Abisko are ecologically very similar and represent well-known *Salix* diversity hotspots, and our sampling covers all common willow species on which the focal leaf- and bud-galler groups are found. The remaining *Salix* species found in these localities occur only sporadically (*S. pentandra*), have associated leaf and/or bud gallers but are rare (*S. myrtilloides*, *S. arbuscula* and *S. caprea*), or are common but not utilized by the focal galler groups (*S. herbacea* and *S. polaris*). Overall, we sequenced 12 individuals per willow host species for both galler groups; however, due to sequencing failure for one leaf galler from *S. glauca*, the total sample size is 179 individuals, of which 84 are bud gallers and 95 are leaf gallers.

2.2.2 | Salix

We constructed a phylogenetic tree for the eight focal *Salix* host species by generating new RADseq data for three species (*S. myrsinites*, *S. phylicifolia* and *S. glauca*) and then combining those with previously published data for the other five species from Wagner et al. (2020). The leaf samples of the three additional species were collected from Finland and Norway (Table S2) and stored in silica gel until DNA extraction.

2.3 | DNA extraction and sequencing

2.3.1 | Gallers

Prior to DNA extraction, the ovipositors of the females were removed and stored as species vouchers in 99.5% ethanol at −20°C. Total genomic DNA was extracted from the remainder of each specimen and then quantified following protocols described in Michell and Nyman (2021).

Uniquely indexed sequencing libraries with an expected fragment size of 350 bp were prepared from 400 ng of input genomic DNA using the NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England Biolabs) and NEBNext Multiplex Oligos for Illumina kits (96 Unique Dual Index Primer Pairs, Item no. E6440S) following the manufacturer's protocols. The libraries were quantified using a Qubit 3.0 system (ThermoFisher Scientific), visualized on 0.8% agarose gels and then pooled in equimolar ratios. The final pooled library was sequenced at the Finnish Institute for Molecular Medicine (FIMM) on a NovaSeq 6000 (Illumina) platform for 2 × 150 bp using a S1 flowcell split into two sections with a lane divider.

2.3.2 | Salix

We extracted DNA from the new samples using the Qiagen DNeasy Plant Mini Kit following the manufacturer's instructions. After a

quality check, the extracts were sent to Floragenex, where RADseq libraries were prepared following Baird et al. (2008) using the *Pst*I restriction enzyme (for details see Wagner et al., 2018).

2.4 | Bioinformatic processing

2.4.1 | Galler nuclear SNP and mtDNA sequence data sets

Sequencing adapters and low-quality bases were trimmed from the raw sequencing libraries using TRIMMOMATIC version 0.39 (Bolger et al., 2014). We removed the 10 leading and trailing bases and applied a sliding window of 4 bp and a Phred quality cutoff of 20 to remove low-quality bases, and then removed sequence pairs in which at least one of the reads was <75 bp long. The trimmed libraries were then quality checked using FASTQC version 0.10.1 (Andrews, 2010) and compiled using MULTIQC (Ewels et al., 2016).

Due to the differential requirements of subsequent population-genetic analyses, we compiled data sets through two different data tracks determined by the reference genomes (Michell et al., 2021) that reads were mapped against. For analyses of differentiation across hosts within the two galler groups, we used the "Reference = own" data stream, in which leaf-galler reads were mapped against the genome of *Eupontania aestiva* (BioProject ID: PRJNA692828, leaf galler on *S. myrsinifolia*; note that the name *Euura saliciscinereae* was used for this species by Liston et al. (2017), who applied the aforementioned generic nomenclature of Prous et al. (2014)), and bud-galler reads against the genome of *Euura lappo* (BioProject ID: PRJNA692175, bud galler on *S. lapponum*). To obtain a combined data set for phylogenetic reconstruction and to allow direct comparisons across the galler groups, we used the "Reference = bud galler" data stream, in which both leaf- and bud-galler reads were mapped against the genome of *Euura lappo*. In both data streams, read mapping was performed using BWA version 0.7.17 (Burrows-Wheeler Aligner; Li & Durbin, 2009), specifically the bwa mem algorithm, with default parameters. PCR (polymerase chain reaction) and optical duplicates were then identified and marked for exclusion using the MarkDuplicates tool in PICARD version 2.21.4 (Broad Institute, 2019), resulting in a BAM file ready for variant calling.

Sequence variants were called using FREEBAYES-PUHTI version 1.3.1 (https://docs.csc.fi/apps/freebayes/) on the Puhti server of the CSC – IT Center for Science. FREEBAYES-PUHTI is a modified version of the FREEBAYES-PARALLEL program which executes the analysis in an automated process, and works by calling variants over 100,000-bp regions in the reference genome and then merging the calls into a final VCF file. We used the default settings while limiting the overall read depth (–g option) to 10,000. Numbers of raw variant calls were calculated using VT-PEEK version 0.57721 (Tan et al., 2015). To determine the sex of the nine leaf-galler larvae from *S. myrsinites*, we calculated their levels of heterozygosity using the -het option in VCFTOOLS version 0.1.16 (Danecek et al., 2011); because the estimates for the larvae corresponded to values of typical diploid females, they were all included in the subsequent data analyses.

The raw VCF file was filtered over three steps in VCFTOOLS. The first filter applied allowed a maximum missing genotype of 5%, a minor allele count of 3 and required a minimum sequencing Phred quality score of 30. Second, variants were filtered for a minor allele frequency of ≥0.05 and a minimum mean sequencing depth of 10× (cf. Jiang et al., 2019). The third filter was applied to the allele balance (AB > 0.25 & AB < 0.75|AB < 0.01). Finally, to obtain only biallelic SNPs, we removed indels and split multi-allelic sites using VCFALLELICPRIMITIVES from VCFLIB (Garrison, 2016) and thinned the variants to only contain SNPs separated by at least 1000 bp using VCFTOOLS. This resulted in three biallelic SNP VCF files with high call quality from the two data streams.

Based on the same sequencing read outputs, we also assembled the sequences of 13 mitochondrial protein-coding genes and two rRNA genes for each galler individual. This was done by following the MITOFINDER pipeline (Allio et al., 2020) with default parameters. The pipeline first assembles all of the sequencing reads *de novo* with MEGAHIT (Li et al., 2016), and then uses BLAST to identify and annotate mitochondrial genes against a user-supplied reference mitochondrial genome. As a reference, we used the sawfly mitogenome published by Sun et al. (2022) under the names *Pontania* (=*Euura s.l.*) *dolichura* in the original publication and *P. bridgmanii* in GenBank (accession MZ726800.1). Based on comparisons to DNA barcode sequence databases, the mitogenome published by Sun et al. (2022) is in fact likely to instead originate from a free-feeding species of *Nematus* (=*Euura s.l.*), but it is still closely related enough to be useful as a reference when assembling mtDNA sequences of galling sawflies.

2.4.2 | Salix

The quality of the 100-bp single-end sequence reads was checked using FASTQC. The reads of all samples were then demultiplexed and analysed in IPYRAD version 0.9.14 (Eaton & Overcast, 2020) with a clustering threshold of 85% and a minimum depth of eight reads for base calling. The maximum number of SNPs per locus was set to 10 and the maximum number of indels to 8. We set a threshold of maximal four alleles per site in the final cluster filtering, because the focal set of species includes several polyploids. IPYRAD summarizes the underlying allelic information into a consensus sequence with ambiguous sites at heterozygous positions. The clustering settings were optimized as described in Wagner et al. (2018), and set to a minimum of six samples sharing a locus (m6). The resultant loci were concatenated for phylogenetic analyses.

2.5 | Estimation of genetic differentiation in gallers

We inferred population structure in the two *Reference = own* data sets using discriminant analysis of principal components (DAPC) in

the POPPR version 2.9.0 (Kamvar et al., 2014) package in R version 4.0.2 (R Development Core Team, 2016) without a priori group assignments. The optimal number of genetic clusters (K_{DAPC}) was determined by the *find.clusters* command with 10,000 iterations in ADEGENET version 2.1.4 (Jombart, 2008; Jombart & Ahmed, 2011), and the K_{DAPC} with the lowest Bayesian information criterion (BIC) was considered optimal. Similarly, the optimal number of principal components to retain in the analysis was determined by the alpha optimization method with the *optim.a.score* function in ADEGENET. Next, we estimated admixture coefficients for each individual in the *Reference = own* data sets based on sparse non-negative matrix factorization (SNMF) in the R package LEA (Frichot & François, 2015). In this case, the number of ancestral populations (K_{SNMF}) was determined by calculating cross-entropy for values of K_{SNMF} from 1 to 19, with 10 replicates for each value; the value with the lowest cross-entropy was considered optimal and was used to calculate the ancestry coefficients (Q matrix). Plotting of the Q matrix was done using GGPLOT2 version 3.3.2 (Ginestet, 2011) in R for the optimal K_{SNMF} and values ±1 of the optimum. The admixture of individuals in the combined *Reference = bud galler* data set was estimated with SNMF in the same manner, except that the optimal K_{SNMF} was determined by evaluating values ranging from 1 to 25.

For the mtDNA data set, we estimated the effect of gall type and willow host species on the distribution of genetic variation through hierarchical analysis of molecular variance (AMOVA) performed in ARLEQUIN version 3.5.2.2. (Excoffier & Lischer, 2010). Before the analysis, we used MODELFINDER (Kalyaanamoorthy et al., 2017) to determine the best-fitting substitution model (=Tamura–Nei) of those implemented in ARLEQUIN, as well as to estimate the alpha shape parameter (=0.181) of the gamma distribution modelling rate heterogeneity across sites of the alignment. Separate analyses were used to estimate Φ_{ST} values across population samples reared from different willow hosts within leaf and bud gallers. In all three analyses, the site-specific maximum proportion of missing data was set to 0.05, and the statistical significance of parameter estimates was estimated through 10,000 randomizations.

2.6 | Phylogenetic reconstruction

2.6.1 | Galler nuclear SNP and mtDNA trees

For the SNP data set, we estimated a maximum-likelihood (ML) phylogeny for all included leaf- and bud-galler individuals in RAXML version 8.2.12 (Stamatakis, 2014) based on the combined *Reference = bud galler* data set and using the GTR+Γ+ASC model as determined by MODELFINDER. The tree was midpoint rooted between the two focal galler groups following the results of Nyman et al. (2007) and Liston et al. (2017), and rapid bootstrapping with 200 replicates was used to infer clade support.

A corresponding individual-level ML tree based on concatenated coding mtDNA gene sequences was constructed by implementing an unpartitioned TPM2u+F+I+G4 model of substitution in IQ-TREE version 1.6.12 (Nguyen et al., 2015), where the ideal substitution model was determined automatically based on the BIC by MODELFINDER. Clade support was estimated based on 1000 bootstrap replications. As a further confirmation of the rooting, we estimated the position of the root using ROOTDIGGER version 1.7.0 (Bettisworth & Stamatakis, 2021) in exhaustive mode.

2.6.2 | Salix

We inferred phylogenetic relationships among the *Salix* host species based on the concatenated alignment of RADseq loci. The ML search in RAXML employed a GTR+Γ model of nucleotide substitution. The tree was rooted by using *S. reticulata* as an outgroup, following the results of Wagner et al. (2018, 2020), and clade support was estimated using rapid bootstrapping with 100 replicates.

2.7 | Cophylogenetic analyses and comparisons across gall types

We used Procrustean Approach to Cophylogeny (PACo; Balbuena et al., 2013; Hutchinson et al., 2017) to estimate the overall congruence between the leaf- and bud-galler SNP phylogenies and that of their *Salix* hosts, as well as to test for congruence between the two galler trees. PACo measures the global congruence between phylogenetic trees through Procrustes fitting of interspecific distance matrices estimated from the trees (Balbuena et al., 2013). Importantly for the present study, PACo allows parasites to be oligophagous (associated with multiple hosts) as well as testing of both symmetric and asymmetric dependencies across phylogenetic trees (Balbuena et al., 2013; Dismukes et al., 2022). To prepare the inputs for the PACo analysis, we first split the overall galler ML tree (see Figure 3) into two trees according to gall types. The gall type-specific trees were then pruned to include only a single representative from each inferred galler species using the keep.tip function in PHANGORN (Schliep, 2011). We note that, because conspecific individuals formed monophyletic clades on the trees, the selection of individuals does not affect the topology of the pruned trees, and the effect on estimated branch lengths will be small. Next, the two galler trees and the *Salix* tree were ultrametricized using penalized likelihood (Sanderson, 2002) in the APE package (Paradis & Schliep, 2019) in R; for this, we used the chronos function with a correlated rate model, with the smoothing parameter lambda and the root age set to 1. The three ultrametric trees (see Figure 4a–c) were converted into distance matrices with the cophenetic function of the STATS package in R. PACo measures phylogenetic congruence based on a residual sum-of-squares value (m^2_{XY}), the significance of which is assessed by 10,000 random permutations of the association matrix. The null hypothesis of no congruence between the two phylogenies is tested by determining the probability that randomly permuted matrices result in an m^2_{XY} that is smaller than the observed value. The *Salix*–galler PACo analyses were based on the symmetrical Procrustes

statistic and the "r0" null model from VEGAN (Oksanen et al., 2019), under the assumption that the phylogeny of the sawflies tracks the evolution of their *Salix* hosts, while the comparison between leaf and bud gallers was based on the "backtracking and swaps" null model (see Hutchinson et al., 2017).

3 | RESULTS

3.1 | Data description

3.1.1 | Galler nuclear SNP and mtDNA sequence data sets

Sequencing on the Illumina NovaSeq resulted in 669.1 Gb of data, resulting in an estimated average of 15× coverage of the genomes per sample, assuming a genome size of 280 Mb (Michell et al., 2021; Oeyen et al., 2020). The sequencing was also of high quality, with 91.1% of the reads having a Phred Q value > 30.

The mapping rate of the leaf-galler samples against the *Eupontania aestiva* genome in the *Reference = own* data track ranged from 91.3% to 98.5%. Variant calling with FREEBAYES identified 20,739,027 raw variants; after applying the above filters, the final variant call file contained 35,629 biallelic SNPs. The mapping rate of the bud-galler samples to the *Euura lappo* genome in the corresponding *Reference = own* data track ranged from 90.1% to 98.1%. Variant calling on these samples resulted in 9,361,780 raw variant calls. After applying the aforementioned filters, the final VCF contained 25,991 biallelic SNPs.

The mapping rate of the leaf-galler samples in the *Reference = bud galler* data set ranged from 88.9% to 96.1%, with an average mapping rate of 95.3%. FREEBAYES identified 22,760,210 variants across all 179 individuals, of which 115,067 high-quality biallelic SNPs remained after filtering. Of these, 29,351 were variable within bud gallers, 65,472 within leaf gallers and 10,229 in both.

The mtDNA alignment comprised sequences of 13 protein-coding genes and two rRNA genes for 179 galler individuals. The average number of assembled and annotated mitochondrial sequences for individual bud and leaf gallers was 13 and 14, respectively. The overall length of the concatenated alignment was 13,397 bp. After filtering out sites with over 5% missing data, the data sets used in ARLEQUIN included 7611 sites for the combined analysis and 9800 and 5391 sites for the separate analyses of leaf and bud gallers, respectively.

3.1.2 | Salix

RADseq of three new species resulted in an average of 6.84 million raw reads per sample. After processing in IPYRAD, the final alignment of all eight *Salix* species was built from 43,182 RADseq loci that comprised 3,734,409 bp and 245,304 SNPs. The alignment contained 11.08% missing data.

3.2 | Patterns of HAD in leaf and bud gallers

The optimal value of K_{DAPC} as determined by the value with the lowest BIC in the two *Reference = own* data sets was eight and five for the leaf and bud gallers, respectively. Alpha optimization showed that four principal components should be retained for the leaf gallers and 11 principal components for the bud gallers in their respective DAPC analyses (Figure S1). The DAPC ordination consistently clustered leaf-galler individuals according to the *Salix* host species that they had been collected from (Figure 2a). The situation in the five bud-galler clusters was more complex, so that three separate groups were formed by individuals collected from *Salix hastata*, *S. lapponum* and *S. lanata*, while two intermixed clusters consisted of individuals from *S. glauca* and *S. myrsinites*, and *S. myrsinifolia* and *S. phylicifolia*, respectively (Figure 2b).

SNMF analyses of the two *Reference = own* data sets revealed essentially similar results. Based on cross-entropy values, the optimal K_{SNMF} was eight for the leaf gallers and five for the bud gallers (Figure S2A,B). In the leaf gallers, very little admixture was present, and >90% of the ancestry coefficient of each individual arose from the specific cluster determined by the host species (Figure S3). At the optimal $K_{SNMF} = 5$, the ancestry coefficient plots of bud gallers showed the same groups as the DAPC ordination, with little admixture being present among clusters (Figure S4). However, one bud-galler individual (BG050) collected from *S. myrsinifolia* appeared to represent an outlier with a high level of admixture. Clustering at $K_{SNMF} = 4$ combined individuals from *S. lanata* with those forming the *S. glauca* + *S. myrsinites* cluster at $K_{SNMF} = 5$. At $K_{SNMF} = 6$, the *S. glauca* + *S. myrsinites* cluster was broken up, but with individual assignments becoming very uncertain (Figure S4).

The combined ML phylogeny based on the *Reference = bud galler* data set grouped the 95 leaf-galler individuals into eight monophyletic groups that were defined by *Salix* host species (Figure 3). Each host-based clade was supported by a 100% bootstrap value, while groupings of individuals within these clades were generally weakly supported. The situation was again more complex within the bud-galler clade of the phylogeny: individuals collected from *S. lapponum*, *S. lanata* and *S. hastata* formed strongly supported monophyletic groups defined by their respective host species. By contrast, individuals reared from *S. glauca* and *S. myrsinites* formed an intermixed group that was paraphyletic with respect to the *S. lanata*-associated clade, and most individuals from *S. myrsinifolia* and *S. phylicifolia* were intermixed with each other. However, the aforementioned individual BG050 reared from *S. myrsinifolia* was placed as sister to a clade formed by the latter group and individuals from *S. hastata*, although this placement was weakly supported.

The SNMF analysis of the joint *Reference = bud galler* data set gave an optimal K_{SNMF} value of 11, but the difference in the level of cross-entropy between a $K_{SNMF} = 11$ and 12 was small (Figure S2C). The phylogenetic clustering of the leaf gallers was confirmed under both values of K_{SNMF}, with eight distinct host-based genetic clusters with very little admixture observed. The number of genetic clusters

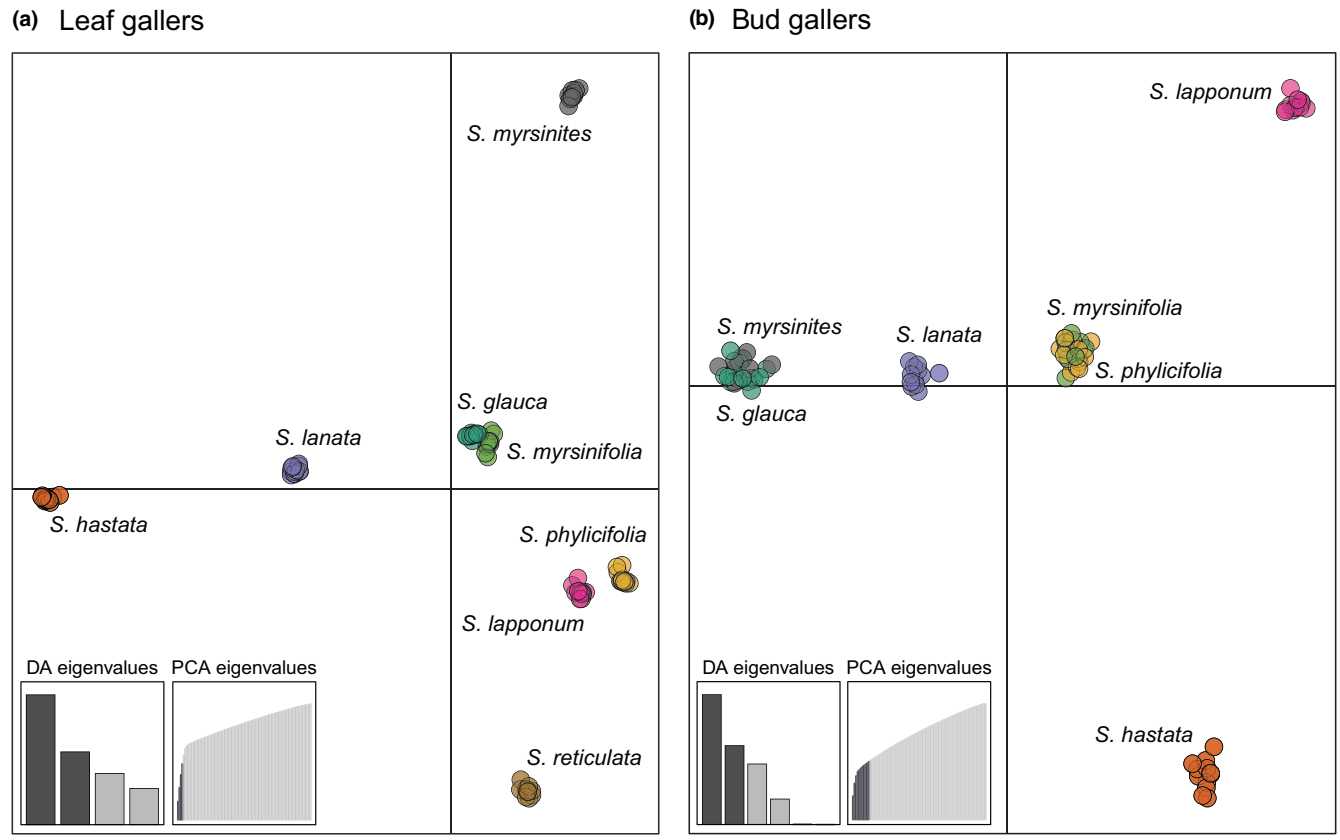


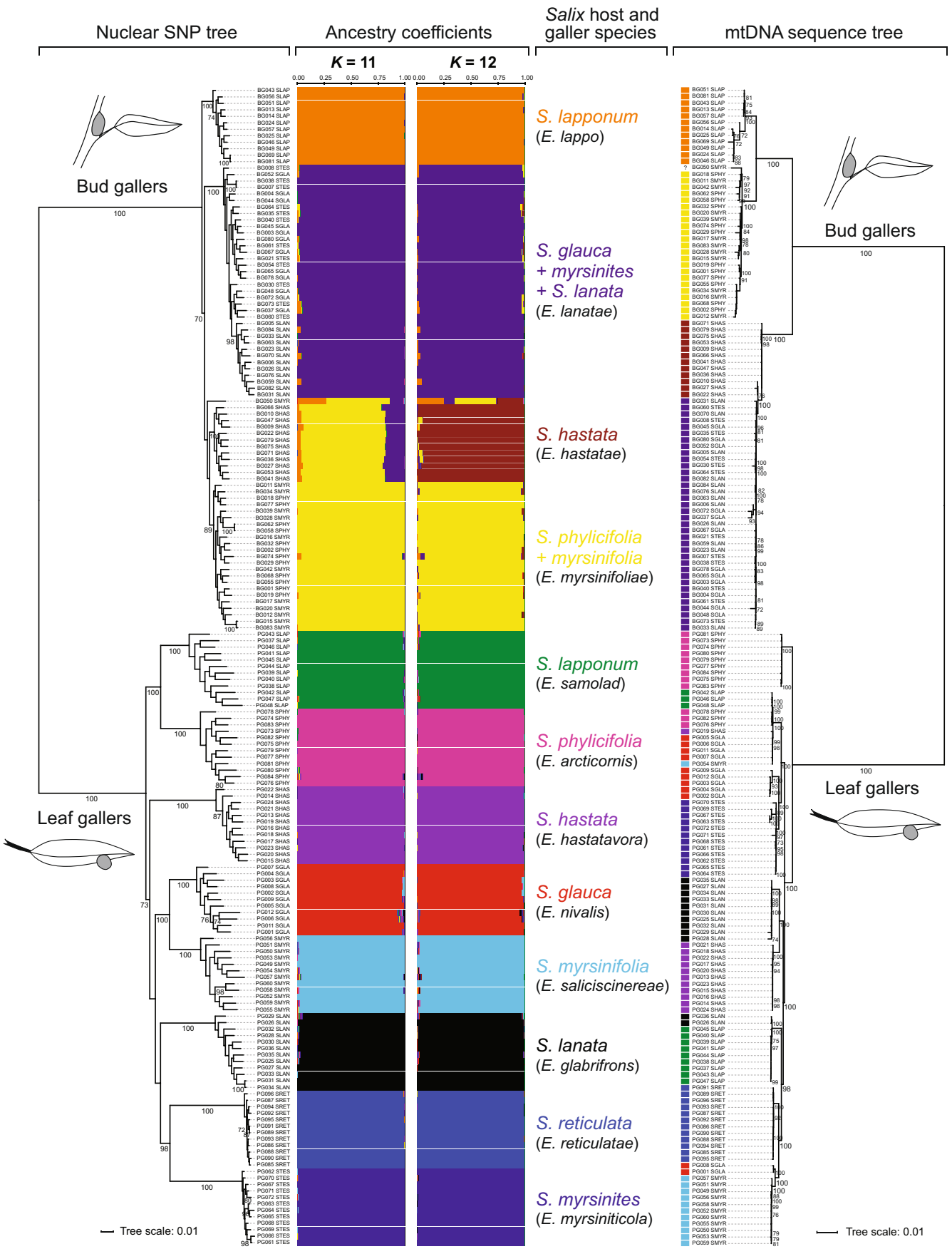
FIGURE 2 DAPC plots for (a) leaf and (b) bud gallers. Dots represent single galler individuals, while dot colours and labels refer to the *Salix* host that each individual sawfly was collected from

in the bud gallers was either three or four depending on the overall K_{SNMF} value (Figure 3). The three genetic clusters identified when K_{SNMF} = 11 were a single distinct cluster for gallers collected on *S. lapponum*, and two clusters consisting of gallers collected on *S. glauca*+*S. myrsinites*+*S. lanata* and *S. hastata*+*S. myrsinifolia*+*S. phylicifolia* (Figure 3). However, at K_{SNMF} = 12, individuals collected on *S. hastata* formed a single genetic cluster separate from the *S. myrsinifolia*+*S. phylicifolia* cluster, with little admixture present (Figure 3).

The phylogenetic tree calculated based on mtDNA sequences was generally strongly supported, with ROOTDIGGER placing the root in the expected position between the two galler groups (Figure 3). In the hierarchical AMOVA based on combined mtDNA data, gall type explained 91.48% and willow host species 7.61% of the overall variation, while within-host variation accounted for 0.91% (all $p < .0001$; Table S3).

Within the bud-galler clade of the mtDNA tree, the main groupings corresponded closely with the results of the SNP-based analysis (Figure 3). Therefore, the first group consisted exclusively of individuals reared from *S. lapponum* and the second one of individuals from *S. phylicifolia* and *S. myrsinifolia*. The third main clade had a nested structure, so that individuals reared from *S. hastata* formed a paraphyletic group with respect to a clade formed by individuals from *S. lanata*, *S. glauca*, and *S. myrsinites*. In the separate analysis of bud-galler haplotypes, estimates of Φ_{ST} across population pairs collected from different hosts were generally high, with the exception of *S. phylicifolia*+*S. myrsinifolia* (Φ_{ST} = 0.033) and pairs involving *S. lanata*, *S. glauca* and *S. myrsinites* (Φ_{ST} = 0.054–0.117; Table S4). Estimates of Φ_{ST} were significantly different from zero across all pairs except *S. phylicifolia*+*S. myrsinifolia* ($p = .198$; Table S5). However, after Bonferroni correction based on the number of tests, also the differentiation between *S. lanata*+*S. myrsinites* as well as *S. glauca*+*S. myrsinites* became nonsignificant.

FIGURE 3 Phylogenetic relationships among the analysed bud- and leaf-galling sawfly individuals and individual-level ancestry coefficients. The two maximum-likelihood phylogenies were calculated based on the combined *Reference = bud galler* nuclear SNP data set (left) and mitochondrial DNA sequences spanning 13 protein-coding and two rRNA genes (right). Names at tips show five-digit individual codes and *Salix* host species with four-letter codes, and numbers below branches are bootstrap proportions (only values >70% shown). Admixture plots for K_{SNMF} = 11 and 12 are shown between the trees; the proportions of bar sections show the admixture coefficients. Squares in front of specimen codes in the mtDNA tree are coloured according to the assignments of the K_{SNMF} = 12 analysis. Names of *Salix* host species (in coloured font) and the inferred galler species (in parentheses, following the nomenclature used by Liston et al., 2017) are shown to the right of the ADMIXTURE plots



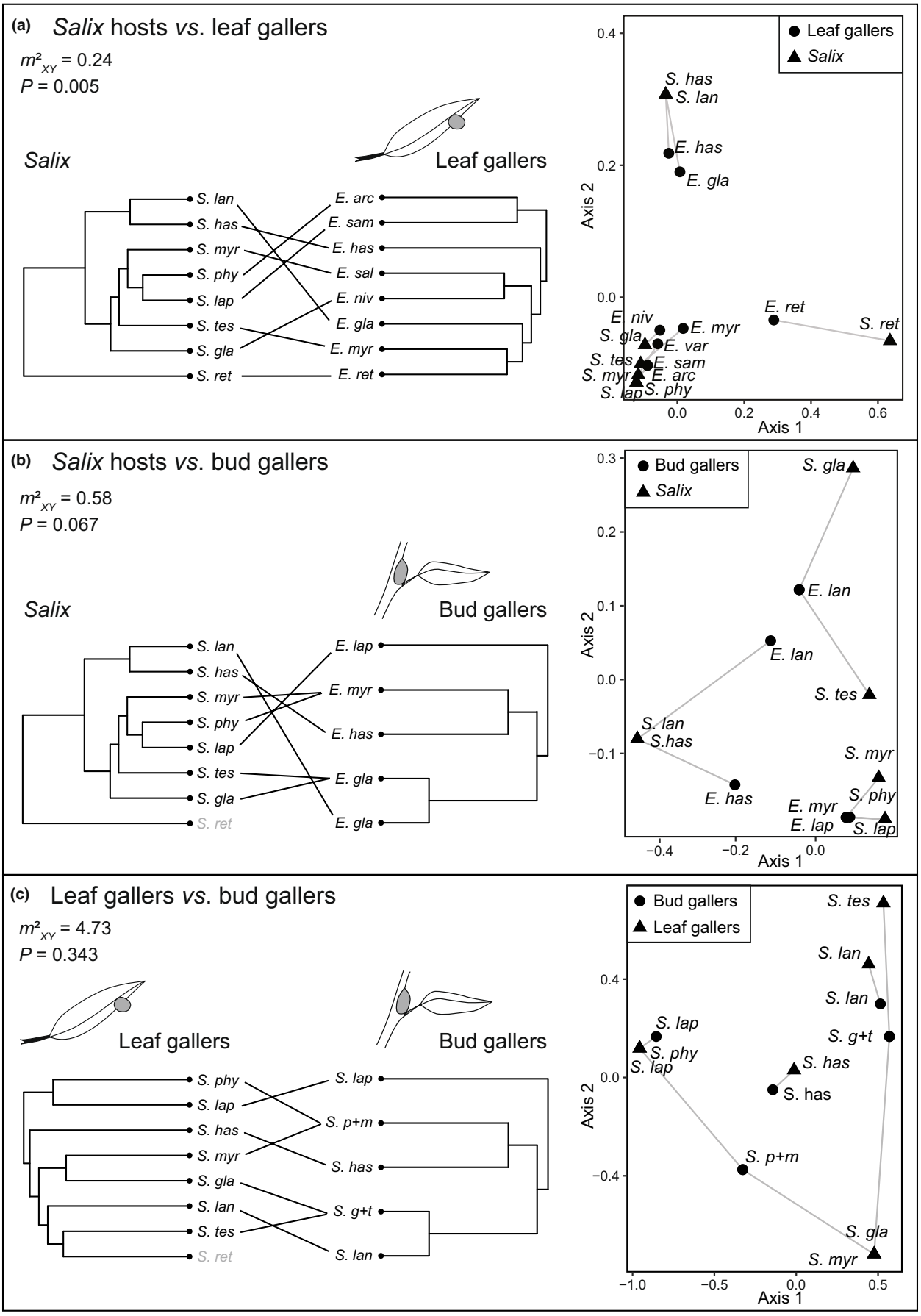


FIGURE 4 Comparisons of phylogenetic trees and corresponding PACo ordination plots for (a) leaf gallers vs. *Salix*, (b) bud gallers vs. *Salix* and (c) leaf gallers vs. bud gallers. Lines connecting species across phylogenies show associations in the *Salix*–galler comparisons and denote correspondences across the two groups in the leaf galler – bud galler plots. Host abbreviations are: *S. gla*, *Salix glauca*; *S. has*, *S. hastata*; *S. lan*, *S. lanata*; *S. lap*, *S. lapponum*; *S. myr*, *S. myrsinifolia*; *S. tes*, *S. myrsinites*; *S. phy*, *S. phylicifolia*; *S. ret*, *S. reticulata*. Galler species names are abbreviated to include only the three first letters (see Figure 3 for full names). Note that in (c), galler tips are labelled with the corresponding *Salix* species codes to facilitate comparisons of host use, and that the *S. ret* tip in (b) and (c) was pruned from the willow and leaf-galler trees used in the PACo analyses involving bud gallers, which do not utilize *S. reticulata*

Within the leaf-galler clade, equally clear monophyletic host-based mtDNA clusters were for the most part not formed, so that only individuals reared from *S. myrsinites* and *S. reticulata* were grouped in their respective exclusive clades (Figure 3). For the samples reared from the six other *Salix* species, 5–11 individuals were grouped in clades defined by each host, while the remaining one to seven individuals were included in groups containing also specimens originating from other *Salix* species. Nevertheless, identical haplotypes were very rarely found across host-based groups, and all pairwise estimates of differentiation were high (Φ_{ST} = 0.357–0.944) and statistically highly significantly different from zero (all $p \leq .001$; Tables S6 and S7).

3.3 | Salix phylogeny

The ML analysis based on RADseq data resulted in a well-resolved phylogeny for the eight focal *Salix* host species (Figure S5). Groupings on the tree are well supported, but we note that the location of the hexaploid *S. myrsinifolia* differs from that in the larger phylogenetic trees published by Wagner et al. (2020). However, our addition of three new species (*S. myrsinites*, *S. phylicifolia* and *S. glauca*) makes direct topological comparisons difficult.

3.4 | Cophylogenetic analyses and comparisons across galler groups

The Procrustean analysis of cophylogeny revealed statistically significant cophylogenetic signal between the leaf gallers and their *Salix* hosts (m^2_{XY} = 0.24, p = .005; Figure 4a). By contrast, the cophylogenetic signal between bud gallers and *Salix* was only marginally significant (m^2_{XY} = 0.58, p = .067; Figure 4b). The correspondence between the leaf- and bud-galler trees was likewise statistically non-significant (m^2_{XY} = 4.72, p = .343; Figure 4c).

4 | DISCUSSION

The repeatability of evolutionary change—on both micro- and macroevolutionary timescales—constitutes a central question in evolutionary biology (Blount et al., 2018; Bohutínská et al., 2021; Bolnick et al., 2018; Gould, 1989; Ord & Summers, 2015). Comparative analyses of parasitic lineages that have radiated across a common resource base can reveal general rules that govern processes of speciation and evolution in the short and long term. Short-term questions concern especially whether independently evolving consumer lineages experience their selective niche landscape in a similar way, that is whether boundaries delimiting host races or species form to encompass the same subsets of available host taxa or resources (Bell et al., 2018; Egan et al., 2013; Johnson et al., 2012; Medina et al., 2017; Stireman 3rd et al., 2005). Questions concerning evolutionary repeatability over longer timescales can be addressed with phylogenetic comparisons, as deterministic processes in radiations driven by sequential host shifts would be expected to result in phylogenetically congruent patterns across replicate radiations (Hamerlinck et al., 2016; Sweet et al., 2016). In both cases, possible discrepancies in host-use patterns and phylogenies across parasitic lineages can be used to pinpoint external abiotic or biotic factors or intrinsic consumer traits that influence patterns of host use and HAD (Dickey & Medina, 2010; Itami & Craig, 2008), and thereby change evolutionary trajectories over phylogenetic timescales (Jousselin et al., 2013; Sweet et al., 2016).

In this study, we compared patterns of HAD as well as long-term cophylogenetic patterns in two related groups of gall-inducing sawflies that have diversified in parallel across northern *Salix* species. While the focal willow–galler system has long been recognized as a promising target for comparative co-evolutionary studies, previous studies applying traditional genetic markers have been hampered by challenges concerning species separation within groups of closely related gallers as well as phylogenetic inference within *Salix*. With the increased resolution afforded by whole-genome resequencing and RADseq markers, we were able to delineate both leaf- and bud-galling sawflies into clear genetic clusters, relate the clusters to information on the hosts that the specimens had been sampled from and, finally, test the level of cophylogenetic congruence between the gallers and their *Salix* hosts as well as between the two galler groups. We demonstrate that, while both leaf- and bud-galling sawflies exhibit specialized host use, the specific patterns of HAD as well as phylogenetic structures differ across the groups. Below, we discuss these differences and their potential causes in detail, and outline directions for future comparative research in species-rich plant–insect systems.

4.1 | Both leaf and bud gallers exhibit HAD, but in different ways

Gall-inducing insects tend to be more specialized in their use of available host plants compared with insect herbivores belonging

to other guilds (Butterill & Novotny, 2015; Oliveira et al., 2020; Stone et al., 2009; Volf et al., 2017). In the galling sawfly subtribe Euurina, mono- or paraphyletic groups inducing different galls on either leaves, buds or shoots have radiated as sequential but temporally and spatially overlapping waves across *Salix* species (Nyman et al., 2000, 2007; Schmidt et al., 2017). The fact that the same willow host species are found in multiple places across the phylogeny of Euurina sawflies in itself indicates that host-shifting must be a frequent phenomenon in co-occurring galler lineages (cf. Ward et al., 2022). Overlapping use of hosts is particularly evident across leaf and bud gallers, which are often found on the same sets of willow species in the same localities (Kopelke, 1999; Kopelke et al., 2017). Unfortunately, the morphological uniformity of willow-galling sawfly species has led to long-lasting debates with regard to the number of species and their host-plant associations (Kopelke, 1999; Liston et al., 2017; Malaise, 1920). At one end of the spectrum is the strict specialization scenario of Kopelke (1999, 2001, 2003), which is based on the view that each galler species is monophagous on a single willow species. On the other hand, for example Vikberg and Zinovjev (2006) and Liston et al. (2017) have postulated that many of the described species cannot be separated based on morphological traits, and therefore have grouped some of the presumed monophages into oligophagous lineages. These uncertainties are reflected in the results of genetic analyses, which have been equally inconclusive: although both Nyman (2002; based on allozymes) and Leppänen et al. (2014; based on mitochondrial COI barcodes and nuclear ITS2 sequences) found clear differences in allele and haplotype frequencies across leaf- and bud-galler samples collected from different willow host species, most differences were not fixed across groups (see also Schmidt et al., 2017). This lack of diagnostic differences meant that the possibility of oligophagous species with differing preferences but partly overlapping host ranges could not be excluded as an explanation for the observed genetic patterns.

By contrast, our ordination, assignment and phylogenetic analyses based on genome-level nuclear SNP and mtDNA sequence data clustered individuals within both leaf and bud gallers into clearly defined groups, to which nearly every resequenced individual can be assigned. In the leaf gallers, each of the eight SNP-based groups is formed exclusively by individuals associated with a single *Salix* species, which supports Kopelke's (1999, 2003) extreme specialization scenario. By contrast, bud gallers evidently represent a mixture of strict specialists and oligophages. In this case, species limits closely match the definitions of Liston et al. (2017), so that two monophagous species occur on *Salix lapponum* (*Euura lappo*) and *S. hastata* (*E. hastatae*), respectively, while two oligophagous species are associated with *S. myrsinifolia* and *S. phylicifolia* (*E. myrsinifoliae*), and *S. lanata*, *S. glauca* and *S. myrsinites* (*E. lanatae*), respectively, though Liston et al. (2017) did not assign specimens from *S. myrsinites* to any species with certainty. Within this last group, individuals reared from *S. lanata* were in some analyses separated from the cluster formed by individuals originating from *S. glauca* and *S. myrsinites*; this partial differentiation could be explained by size-dependent assortative mating (which is a common phenomenon in insects; Jiang et al., 2013), as galls on the relatively robust *S. lanata* tend to be big and, hence, support large galler larvae that emerge as large adults (Liston et al., 2017).

Analyses of mtDNA sequence variation largely confirmed the SNP-based results, despite differences in details. Within bud gallers, the main mtDNA clusters corresponded directly with the limits of the SNP-based delimitations, although the backbone structures of the nuclear and mtDNA trees were in conflict. In leaf gallers, however, we instead found clear frequency differences across *Salix* host species but an absence of strict reciprocal monophyly, a pattern resembling the barcode-based results of Leppänen et al. (2014). Mitonuclear discordance following from incomplete lineage sorting or hybridization is common in insects (e.g., Campbell et al., 2022; Linnen & Farrell, 2007; Lopez-Vaamonde et al., 2021), and the haplodiploid sex determination system of sawflies and other hymenopterans biases introgression towards mitochondria (Patten et al., 2015; Prous et al., 2020). Porous species boundaries are a probable explanation for the genetic composition of bud galler BG050 in our data set: the specimen was reared from *S. myrsinifolia* and was grouped within the *E. myrsinifoliae* clade in the mtDNA tree, but appears to represent a case of mixed ancestry based on SNP data (Figure 3).

4.2 | Leaf galler and Salix phylogenies show congruence, but leaf galler and bud galler phylogenies do not

Comparative phylogenetic studies have shown that parasites can occasionally cospeciate along with their hosts (Hamerlinck et al., 2016; Sweet et al., 2016), but that strict long-term parallel cladogenesis is rare in species-rich systems (Brown et al., 2022; Cruaud et al., 2012; de Vienne et al., 2013; Nylin et al., 2018). This is true also for most plant–insect networks, in which phylogenetic discordance arises because of occasional host shifting by herbivores (Hsu et al., 2018; Suchan & Alvarez, 2015; Ward et al., 2022). On the other hand, the likelihood of cospeciation increases whenever associates are strictly specialized to single host species (Banks & Paterson, 2005; but see Hayward et al., 2021), which is the case in many gall-inducing insect taxa (Hardy & Cook, 2010). As mentioned above, the fact that many willow species are utilized by gallers representing several Euurina galler groups in itself is evidence against long-term parallel cladogenesis between sawfly gallers and their *Salix* hosts (Nyman et al., 2000, 2007; see also Liston et al., 2017). Available evidence also indicates that the *Salix*–galler system conforms to the usual pattern in plant–insect networks, with host-plant taxa being considerably older than their associated herbivores (cf. Leppänen et al., 2012; Lopez-Vaamonde et al., 2006; McKenna et al., 2009; Percy et al., 2004). The oldest willow fossils originate from the early Eocene in North America (Collinson, 1992), and Wu et al. (2015) estimated the age of the crown node of *Salix* to be *c*. 43 million years ago (Ma). Given that willow gallers are nested within the Nematini tribe of the tenthredinid sawfly subfamily Nematinae (Nyman et al., 2010; Prous

et al., 2014), the fossil-calibrated Hymenoptera phylogeny of Nyman et al. (2019) suggests an age of less than 10 Ma for the willow gallers. However, the main radiation of *Salix* subgenus *Vetrix*, which is the most species-rich willow subgenus and includes the species that host the majority of willow-galling sawflies, commenced *c*. 20 Ma (Wu et al., 2015), with a possible acceleration occurring during the Pleistocene glaciations (Lauron-Moreau et al., 2015). This could make the main radiation of subgenus *Vetrix* roughly contemporaneous with those of the leaf- and bud-galling clades studied here, as these taxa are phylogenetically nested inside Euurina galler groups inducing leaf rolls and other types of closed galls (Liston et al., 2017; Nyman et al., 2000, 2007).

Although leaf- and bud-galling sawflies have probably diversified after the origin and radiation of *Salix*, a hypothesis of evolutionary repeatability would still predict some level of phylogenetic congruence across insect lineages that are independently diversifying by host-shifting among species within the same host-plant taxon (cf. Hayward et al., 2021; Sweet et al., 2016). Against the backdrop of generally discordant patterns between galler and willow phylogenies, it was somewhat surprising that the Procrustean ordination analysis indicated a statistically significant level of phylogenetic congruence between leaf gallers and their *Salix* hosts. Nevertheless, the congruence between the two phylogenies is far from perfect, and the cophylogenetic signal is driven primarily by a few links with low jackknifed residuals (Figure 4a). The result may therefore be a false positive resulting from preferential shifting among related host species, which tend to share ecologically relevant traits (de Vienne et al., 2007; Futuyma & Agrawal, 2009). For the bud gallers, the same analysis indicated no congruence with the willow phylogeny, which could partly result from low statistical power as a result of the lower number of tips on the bud-galler tree. On the other hand, it is noteworthy that the *Salix* host species of the two oligophagous bud-galler species do not constitute monophyletic groups (Figure 4b). Hence, the delimitation of host repertoires in these bud-galling species (as a result of host selection by females and successful gall induction) is evidently based on traits that are not directly congruent with the *Salix* phylogeny (cf. Clayton et al., 2003).

Explaining the host ranges of the two oligophagous bud-galler lineages is far from straightforward using the "standard" explanation of chemical similarity among plant taxa. Plant chemistry often directs species-level host use (Volf et al., 2015) and can lead to long-term resource tracking (Becerra & Venable, 1999; Endara et al., 2018; Murphy & Feeny, 2006) in plant-feeding insects. When considering leaf chemistry, *S. myrsinifolia* contains high levels of diverse phenolic glycosides, while *S. phylicifolia* is generally considered to be mildly defended (Julkunen-Tiitto, 1989; Nyman & Julkunen-Tiitto, 2005). Likewise, *S. myrsinites* and *S. glauca* are chemically relatively similar and strongly defended, but *S. lanata* is not (Julkunen-Tiitto, 1989; Nyman & Julkunen-Tiitto, 2005). Importantly, however, the *S. lanata*+*S. myrsinites*+*S. glauca* triplet is more homogeneous when considering the chemistry of shoots (Julkunen-Tiitto, 1989), on which bud-galler females find the developing buds that they oviposit into (Kopelke, 1999). Nevertheless, the available data suggest that *S. phylicifolia* and *S. myrsinifolia* differ also in their shoot chemistry (Julkunen-Tiitto, 1989).

An alternative possibility is that the host ranges of the two oligophagous species represent long-term legacies of overlapping distributions and habitat of their *Salix* hosts. While our galler samples originated from northern Fennoscandia, the geographical distributions of arctic plant and insect communities are known to have undergone extensive latitudinal and longitudinal migrations throughout the Pleistocene (Brochmann et al., 2013; Eidesen et al., 2013), and our focal study areas were covered by the Scandinavian ice sheet until <10,000 years ago (Regnéll et al., 2019). As shown by McBride et al. (2009) and Linnen and Farrell (2010), host races and differentially specialized sister species of insect herbivores that are sympatric today may have originated in allopatry (see also Pérez-Pereira et al., 2017). Furthermore, Calatayud et al. (2016) demonstrated that long-term patterns of host shifts in plant-associated spider mites have been affected by past geographical proximity (i.e., availability for colonization). In our focal willow–galler system, the geographical distributions of *S. phylicifolia* and *S. myrsinifolia* are generally more southern than for those of the arctic–alpine *S. lanata*, *S. glauca* and *S. myrsinites* (Skvortsov, 1999). However, with respect to preferred habitat, the latter three species overlap extensively with *S. hastata* and *S. lapponum*, which host their own specialist bud-galling species (Figures 2 and 3).

Interestingly, the different levels of phylogenetic congruence of leaf and bud gallers in relation to their shared *Salix* hosts leads to the situation that the phylogenies of the two focal galler groups are in conflict with each other (Figure 4c). An additional level of conflict is added by the fact that leaf gallers have been able to colonize the creeping tundra willow *S. reticulata*, which is not utilized by bud gallers (Figures 1 and 4). These discordances directly indicate that long-term diversification patterns have been different in the two groups, despite the fact that they are closely related and have radiated roughly synchronously across a shared resource base. A few previous comparative analyses involving herbivore–herbivore comparisons have revealed conflicting phylogenetic patterns across independently evolving pollinator and/or herbivore lineages (Marussich & Machado, 2007; Mlynarek & Heard, 2018). Discordant phylogenies across replicate parasite radiations on the same hosts have also been observed in other host–parasite systems, including cestode and nematode helminths of pikas (Galbreath & Hoberg, 2015), wing and body lice on doves (Sweet et al., 2016), and pinworms of chipmunks (Bell et al., 2018).

4.3 | Conclusions and future directions

Our study based on genome-wide SNP markers and mtDNA sequence data showed the presence of specialized host use and HAD in both leaf- and bud-galling sawflies across a set of shared *Salix* host species, but also revealed considerable discordances in short- and long-term evolutionary patterns between the two focal groups. In a few previous comparative analyses of insect herbivores

and other parasitic taxa, differences in specialization and long-term host use have been linked to candidate traits such as differences in breeding systems (Dickey & Medina, 2010) or capacity for dispersal among host individuals (Sweet et al., 2016). Our analysis differs from previous works in that our focal taxa are closely related and, therefore, very similar in their general biology and ecological traits. Further work is therefore needed to reveal the factors underlying the discrepancies in host use, but they could, for example, be related to differential patterns of chemical or morphological similarity across hosts at the sites that the two galler groups use for oviposition (petiole bases for bud gallers and leaf midribs for leaf gallers; Kopelke, 1999). However, the discordant phylogenies could also arise from chance effects, such as nonoverlapping past geographical distributions leading to different sets of hosts being available for colonization for leaf- and bud-galler lineages. Taken together, our findings suggest that the colonization of available hosts, the buildup of HAD as well as long-term patterns of speciation and niche diversification in insect herbivores are largely dictated by clade-level idiosyncrasies and historical contingencies (see also de Medeiros & Farrell, 2020).

Plant–insect networks constitute an essentially untapped resource for studying evolutionary repeatability in radiations driven by niche shifts. It can be argued that all sister clades or contemporaneously radiating parasitic lineages constitute replicates of the same process, especially when they are restricted to specific host taxa (Blount et al., 2018; Braga et al., 2021; Sweet et al., 2016). In the Euurina gallers alone, four to five lineages inducing different galls have coradiated on willow hosts across the Holarctic region (Liston et al., 2017; Nyman et al., 2000, 2007). The present study provides a snapshot of host use in two of these galler groups in a single—albeit unusually species-rich—geographical region. Obtaining a full view of evolutionary repeatability across sawfly galler radiations will require surveys of host-use patterns in multiple galler clades across the over 400 willow species that occur globally (Wu et al., 2015). The ongoing molecular revolution means that such studies will soon be within reach, shifting the research bottleneck from the collection and analysis of genetic data to obtaining samples with relevant biological background information across vast geographical areas (cf. Johnson, 2019). Importantly, genomic data sets used for inferring patterns of niche use can also be used for uncovering the genetic basis of adaptation to alternative hosts (Berner & Salzburger, 2015; Waters & McCulloch, 2021). A major breakthrough is also provided by the emerging understanding of the age and internal phylogeny of *Salix* (Wagner et al., 2018, 2020; Wu et al., 2015), which will in the near future at last allow studies addressing the co-evolutionary interactions between this species-rich and ecologically central plant group and its diverse herbivores.

AUTHOR CONTRIBUTIONS

Tommi Nyman and Craig T. Michell designed the research, Tommi Nyman collected samples, CTM performed laboratory work and bioinformatic processing, Craig T. Michell, Tommi Nyman and Natascha Wagner analysed the data, and Craig T. Michell and Tommi Nyman wrote the paper with significant inputs from all coauthors. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Funding for this study was provided by the Academy of Finland (Project 294466 to TN). Computation-intensive bioinformatic and statistical analyses were performed on the servers of the Finnish Centre for Scientific Computing (www.csc.fi). Jens-Peter Kopelke kindly provided photographs of bud galls on *S. myrsinifolia* and *S. hastata*. The manuscript benefited from discussions with Catherine Linnen and Phillip Watts, as well as from suggestions provided by six anonymous reviewers.

CONFLICT OF INTEREST

The Authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Sawfly voucher samples are stored in the Biobank of the Svanhovd Research Station of the Norwegian Institute of Bioeconomy Research (NIBIO). The raw willow-galler sequencing data used in this research are available on the NCBI Sequence Read Archive (SRA) under BioProject ID: PRJNA751456. The final galler SNP and mtDNA and *Salix* RADseq data sets, as well as scripts and data files used in the statistical analyses have been deposited in the Dryad repository (Michell et al., 2022; https://doi.org/10.5061/dryad.34tmpg4kv).

ORCID

Craig T. Michell https://orcid.org/0000-0003-4706-7256
Natascha Wagner https://orcid.org/0000-0001-6623-7623
Kyung Min Lee https://orcid.org/0000-0001-9055-1073
Tommi Nyman https://orcid.org/0000-0003-2061-0570

REFERENCES

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*, *20*, 892–905.

Althoff, D. M. (2008). A test of host-associated differentiation across the "parasite continuum" in the tri-trophic interaction among yuccas, bogus yucca moths, and parasitoids. *Molecular Ecology*, *17*, 3917–3927.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Argus, G. W. (1997). Infrageneric classification of *Salix* (Salicaceae) in the New World. *Systematic Botany Monographs*, *52*, 1–121.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*, e3376.

Balbuena, J. A., Míguez-Lozano, R., & Blasco-Costa, I. (2013). PACo: A novel Procrustes application to cophylogenetic analysis. *PLoS One*, *8*, e61048.

Banks, J. C., & Paterson, A. M. (2005). Multi-host parasite species in cophylogenetic studies. *International Journal for Parasitology*, *35*, 741–746.

Becerra, J. X., & Venable, D. L. (1999). Macroevolution of insect–plant associations: The relevance of host biogeography to host affiliation. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 12626–12631.

Bell, K. C., Demboski, J. R., & Cook, J. A. (2018). Sympatric parasites have similar host-associated, but asynchronous, patterns of diversification. *The American Naturalist*, *192*, E106–E119.

Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*, 491–499.

Bettisworth, B., & Stamatakis, A. (2021). Root Digger: A root placement program for phylogenetic trees. *BMC Bioinformatics*, *22*, e225.

Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex determination, sex chromosomes, and karyotype evolution in insects. *The Journal of Heredity*, *108*, 78–93.

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science*, *362*, eaam5979.

Bohutínská, M., Vlček, J., Yair, S., Laenen, B., Konečná, V., Fracassetti, M., Slotte, T., & Kolář, F. (2021). Genomic basis of parallel adaptation varies with divergence in Arabidopsis and its relatives. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2022713118.

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

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non)Parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 303–330.

Boyd, B. M., Nguyen, N. P., Allen, J. M., Waterhouse, R. M., Vo, K. B., Sweet, A. D., Clayton, D. H., Bush, S. E., Shapiro, M. D., & Johnson, K. P. (2022). Long-distance dispersal of pigeons and doves generated new ecological opportunities for host-switching and adaptive radiation by their parasites. *Proceedings of the Royal Society B: Biological Sciences*, *289*, e20220042.

Braga, M. P., Janz, N., Nylin, S., Ronquist, F., & Landis, M. J. (2021). Phylogenetic reconstruction of ancestral ecological networks through time for pierid butterflies and their host plants. *Ecology Letters*, *24*, 2134–2145.

Braga, M. P., Landis, M. J., Nylin, S., Janz, N., & Ronquist, F. (2020). Bayesian inference of ancestral host–parasite interactions under a phylogenetic model of host repertoire evolution. *Systematic Biology*, *69*, 1149–1162.

Broad Institute. (2019). *Picard toolkit*. Broad Institute, GitHub repository. https://broadinstitute.github.io/picard/

Brochmann, C., Edwards, M., & Alsos, I. G. (2013). The dynamic past and future of Arctic vascular plants: Climate change, spatial variation and genetic diversity. In K. Rohde (Ed.), *The balance of nature and human impact* (pp. 133–152). Cambridge University Press.

Brooks, D., & McLennan, D. (2002). *The nature of diversity: An evolutionary voyage of discovery*. The University of Chicago Press.

Brown, A. M., Speer, K. A., Teixeira, T., Clare, E., Simmons, N. B., Balbuena, J. A., Dick, C. W., Dittmar, K., & Perkins, S. (2022). Phylogenetic and ecological trends in specialization: Disentangling the drivers of ectoparasite host specificity. *bioRxiv*. https://doi.org/10.1101/2022.04.06.487338

Butterill, P. T., & Novotny, V. (2015). Gall-forming insects in a lowland tropical rainforest: Low species diversity in an extremely specialised guild. *Ecological Entomology*, *40*, 409–419.

Calatayud, J., Hórreo, J. L., Madrigal-González, J., Migeon, A., Rodríguez, M. Á., Magalhães, S., & Hortal, J. (2016). Geography and major host evolutionary transitions shape the resource use of plant parasites. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 9840–9845.

Campbell, E. O., MacDonald, Z. G., Gage, E. V., Gage, R. V., & Sperling, F. A. H. (2022). Genomics and ecological modelling clarify species integrity in a confusing group of butterflies. *Molecular Ecology*, *31*, 2400–2417.

Cerca, J., Maurstad, M. F., Rochette, N. C., Rivera-Colón, A. G., Rayamajhi, N., Catchen, J. M., & Struck, T. H. (2021). Removing the bad apples: A simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. *Methods in Ecology and Evolution*, *12*, 805–817.

Clayton, D. H., Al-Tamimi, S., & Johnson, K. P. (2003). The ecological basis of coevolutionary history. In R. D. M. Page (Ed.), *Tangled trees: Phylogeny, cospeciation and coevolution* (pp. 310–341). University of Chicago Press.

Collinson, M. E. (1992). The early fossil history of Salicaceae: A brief review. *Proceedings of the Royal Society of Edinburgh, Section B: Biological Sciences*, *98*, 155–167.

Cruaud, A., & Rasplus, J. Y. (2016). Testing cospeciation through large-scale cophylogenetic studies. *Current Opinion in Insect Science*, *18*, 53–59.

Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L. S., Clement, W. L., Couloux, A., Cousins, B., Genson, G., Harrison, R. D., Hanson, P. E., Hossaert-Mckey, M., Jabbour-Zahab, R., Jousselin, E., Kerdelhué, C., Kjellberg, F., Lopez-Vaamonde, C., Peebles, J., Peng, Y. Q., Pereira, R. A. S., … Savolainen, V. (2012). An extreme case of plant-insect codiversification: Figs and fig-pollinating wasps. *Systematic Biology*, *61*, 1029–1047.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

de Medeiros, B. A. S., & Farrell, B. D. (2020). Evaluating insect-host interactions as a driver of species divergence in palm flower weevils. *Communications Biology*, *3*, 749.

de Vienne, D. M., Giraud, T., & Shykoff, J. A. (2007). When can host shifts produce congruent host and parasite phylogenies? A simulation approach. *Journal of Evolutionary Biology*, *20*, 1428–1438.

de Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. *The New Phytologist*, *198*, 347–385.

Dickey, A. M., & Medina, R. F. (2010). Testing host-associated differentiation in a quasi-endophage and a parthenogen on native trees. *Journal of Evolutionary Biology*, *23*, 945–956.

Dismukes, W., Braga, M. P., Hembry, D. H., Heath, T. A., & Landis, M. J. (2022). Cophylogenetic methods to untangle the evolutionary history of ecological interactions. *Annual Review of Ecology, Evolution, and Systematics*, *53*, 275–298.

Dorchin, N., Joy, J. B., Hilke, L. K., Wise, M. J., & Abrahamson, W. G. (2015). Taxonomy and phylogeny of the *Asphondylia* species (Diptera: Cecidomyiidae) of North American goldenrods: Challenging morphology, complex host associations, and cryptic speciation. *Zoological Journal of the Linnean Society*, *174*, 265–304.

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*, 471–492.

Drummond, C. S., Xue, H.-J., Yoder, J. B., & Pellmyr, O. (2010). Host-associated divergence and incipient speciation in the yucca moth *Prodoxus coloradensis* (Lepidoptera: Prodoxidae) on three species of host plants. *Heredity*, *105*, 183–196.

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*, 2592–2594.

Egan, S. P., Hood, G. R., DeVela, G., & Ott, J. R. (2013). Parallel patterns of morphological and behavioral variation among host-associated populations of two gall wasp species. *PLoS One*, *8*, e54690.

Eidesen, P. B., Ehrich, D., Bakkestuen, V., Alsos, I. G., Gilg, O., Taberlet, P., & Brochmann, C. (2013). Genetic roadmap of the Arctic: Plant dispersal highways, traffic barriers and capitals of diversity. *The New Phytologist*, *200*, 898–910.

Endara, M.-J., Nicholls, J. A., Coley, P. D., Forrister, D. L., Younkin, G. C., Dexter, K. G., Kidner, C. A., Pennington, R. T., Stone, G. N., & Kursar,

T. A. (2018). Tracking of host defenses and phylogeny during the radiation of Neotropical Inga-feeding sawflies (Hymenoptera; Argidae). *Frontiers in Plant Science*, *9*, e1237.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*, 3047–3048.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K. G., Widmayer, H. A., & Wilson, C. J. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, *71*, 1126–1137.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929.

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 18054–18061.

Galbreath, K. E., & Hoberg, E. P. (2015). Host responses to cycles of climate change shape parasite diversity across North America's Intermountain West. *Folia Zoologica*, *64*, 218–232.

Garrison, E. (2016). *Vcflib, a C++ library for parsing and manipulating VCF files*. GitHub. https://github.com/vcflib/vcflib

Ginestet, C. (2011). ggplot2: Elegant graphics for data analysis. *Journal of the Royal Statistical Society Series A (Statistics in Society)*, *174*, 245–246.

Gould, S. J. (1989). *Wonderful life: The burgess shale and the nature of history*. W.W. Norton New York.

Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J. P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B., Tylianakis, J. M., Wood, S. A., & Roslin, T. (2019). Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, *42*, 401–415.

Hamerlinck, G., Hulbert, D., Hood, G. R., Smith, J. J., & Forbes, A. A. (2016). Histories of host shifts and cospeciation among free-living parasitoids of *Rhagoletis* flies. *Journal of Evolutionary Biology*, *29*, 1766–1779.

Hardy, N. B., & Cook, L. G. (2010). Gall-induction in insects: Evolutionary dead-end or speciation driver? *BMC Evolutionary Biology*, *10*, e257.

Hardy, N. B., Kaczvinsky, C., Bird, G., & Normark, B. B. (2020). What we don't know about diet-breadth evolution in herbivorous insects. *Annual Review of Ecology, Evolution, and Systematics*, *51*, 103–122.

Hay, E. M., Poulin, R., & Jorge, F. (2020). Macroevolutionary dynamics of parasite diversification: A reality check. *Journal of Evolutionary Biology*, *33*, 1758–1769.

Hayward, A., Poulin, R., & Nakagawa, S. (2021). A broadscale analysis of host-symbiont cophylogeny reveals the drivers of phylogenetic congruence. *Ecology Letters*, *24*, 1681–1696.

Hembry, D. H., & Weber, M. G. (2020). Ecological interactions and macroevolution: A new field with old roots. *Annual Review of Ecology, Evolution, and Systematics*, *51*, 215–243.

Hood, G. R., Powell, T. H. Q. Q., Doellman, M. M., Sim, S. B., Glover, M., Yee, W. L., Goughnour, R. B., Mattsson, M., Schwarz, D., & Feder, J. L. (2020). Rapid and repeatable host plant shifts drive reproductive isolation following a recent human-mediated introduction of the apple maggot fly, *Rhagoletis pomonella*. *Evolution*, *74*, 156–168.

Hsu, Y.-H., Cocroft, R. B., Snyder, R. L., & Lin, C.-P. (2018). You stay, but I hop: Host shifting near and far co-dominated the evolution of *Enchenopa treehoppers*. *Ecology and Evolution*, *8*, 1954–1965.

Hutchinson, M. C., Cagua, E. F., Balbuena, J. A., Stouffer, D. B., & Poisot, T. (2017). Paco: Implementing Procrustean approach to cophylogeny in R. *Methods in Ecology and Evolution*, *8*, 932–940.

Itami, J. K., & Craig, T. P. (2008). 2. Evolution of preference and performance relationships. In K. Tilmon (Ed.), *Specialization, speciation, and radiation* (pp. 20–28). University of California Press.

Janz, N., & Nylin, S. (2008). The oscillation hypothesis of host-plant range and speciation. In K. Tilmon (ed.), *Specialization, speciation, and radiation: The evolutionary biology of herbivorous insects* (pp. 203–215). University of California Press.

Jiang, Y., Bolnick, D. I., Kirkpatrick, M., Arnqvist, A. E. G., & McPeek, E. M. A. (2013). Assortative mating in animals. *The American Naturalist*, *181*, E125–E138.

Jiang, Y., Jiang, Y., Wang, S., Zhang, Q., & Ding, X. (2019). Optimal sequencing depth design for whole genome re-sequencing in pigs. *BMC Bioinformatics*, *20*, e556.

Johnson, K. P. (2019). Putting the genome in insect phylogenomics. *Current Opinion in Insect Science*, *36*, 111–117.

Johnson, K. P., Shreve, S. M., & Smith, V. S. (2012). Repeated adaptive divergence of microhabitat specialization in avian feather lice. *BMC Biology*, *10*, e52.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405.

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071.

Jousselin, E., Cruaud, A., Genson, G., Chevenet, F., Foottit, R. G., & Cœur d'acier, A. (2013). Is ecological speciation a major trend in aphids? Insights from a molecular phylogeny of the conifer-feeding genus *Cinara*. *Frontiers in Zoology*, *10*, e56.

Julkunen-Tiitto, R. (1989). Phenolic constituents of *Salix*: A chemotaxonomic survey of further Finnish species. *Phytochemistry*, *28*, 2115–2125.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589.

Kamvar, Z. N., Tabima, J. F., & Gr̈unwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kawecki, T. J. (1998). Red Queen meets Santa Rosalia: Arms races and the evolution of host specialization in organisms with parasitic lifestyles. *The American Naturalist*, *152*, 635–651.

Kopelke, J.-P. (1999). Gallenerzeugende Blattwespen Europas – Taxonomische Grundlagen, Biologie und Ökologie (Tenthredinidae: Nematinae: Euura, Phyllocolpa, Pontania). *Courier Forschungsinstitut Senckenberg*, *212*, 1–183.

Kopelke, J.-P. (2001). Die Artengruppen von *Euura mucronata* und *E. laeta* in Europa (Insecta, Hymenoptera, Thenthredinidae, Nematinae). *Senckenbergiana Biologica*, *81*, 191–225.

Kopelke, J.-P. (2003). Gall-forming Nematinae, their willow hosts (*Salix* spp.) and biological strategies (Insecta, Hymenoptera, Symphyta, Tenthredinidae, Nematinae: Euura, Phyllocolpa, Pontania). *Senckenbergiana Biologica*, *82*, 163–189.

Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S., & Roslin, T. (2017). Food-web structure of willow-galling sawflies and their natural enemies across Europe. *Ecology*, *98*, 1730.

Kuchta, R., Řehulková, E., Francová, K., Scholz, T., Morand, S., & Šimková, A. (2020). Diversity of monogeneans and tapeworms in cypriniform fishes across two continents. *International Journal for Parasitology*, *50*, 771–786.

Lafferty, K. D., DeLeo, G., Briggs, C. J., Dobson, A. P., Gross, T., & Kuris, A. M. (2015). A general consumer-resource population model. *Science*, *349*, 854–857.

Lauron-Moreau, A., Pitre, F. E., Argus, G. W., Labrecque, M., & Brouillet, L. (2015). Phylogenetic relationships of American willows (*Salix* L., Salicaceae). *PLoS One*, *10*, e0121965.

Leppänen, S. A., Altenhofer, E., Liston, A. D., & Nyman, T. (2012). Phylogenetics and evolution of host-plant use in leaf-mining sawflies (Hymenoptera: Tenthredinidae: Heterarthrinae). *Molecular Phylogenetics and Evolution*, *64*, 331–341.

Leppänen, S. A., Malm, T., Värri, K., & Nyman, T. (2014). A comparative analysis of genetic differentiation across six shared willow host species in leaf- and bud-galling sawflies. *PLoS One*, *9*, e116286.

Leskinen, E., & Alström-Rapaport, C. (1999). Molecular phylogeny of Salicaceae and closely related Flacourtiaceae: Evidence from 5.8 S, ITS 1 and ITS 2 of the rDNA. *Plant Systematics and Evolution*, *215*, 209–227.

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., Yamashita, H., & Lam, T. W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, *102*, 3–11.

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

Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. *Evolution*, *61*, 1417–1438.

Linnen, C. R., & Farrell, B. D. (2010). A test of the sympatric host race formation hypothesis in Neodiprion (Hymenoptera: Diprionidae). *Proceedings of the Royal Society B: Biological Sciences*, *277*, 3131–3138.

Liston, A. D., Heibo, E., Prous, M., Vårdal, H., Nyman, T., & Vikberg, V. (2017). North European gall-inducing Euura sawflies (Hymenoptera, Tenthredinidae, Nematinae). *Zootaxa*, *4302*, 1–115.

Lopez-Vaamonde, C., Kirichenko, N., Cama, A., Doorenweerd, C., Godfray, H. C. J., Guiguet, A., Gomboc, S., Huemer, P., Landry, J., Laštůvka, A., Laštůvka, Z., Lee, K., Lees, D., Mutanen, M., van Nieukerken, E. J., Segerer, A., Triberti, P., Wieser, C., & Rougerie, R. (2021). Evaluating DNA barcoding for species identification and discovery in European gracillariid moths. *Frontiers in Ecology and Evolution*, *9*, e626752.

Lopez-Vaamonde, C., Wikström, N., Labandeira, C., Godfray, H. C. J., Goodman, S. J., & Cook, J. M. (2006). Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, *19*, 1314–1326.

Malaise, R. (1920). Beiträge zur Kenntnis schwedischer Blattwespen. *Entomologisk Tidskrift*, *41*, 97–128.

Martinů, J., Sychra, O., Literák, I., Čapek, M., Gustafsson, D. L., & Štefka, J. (2015). Host generalists and specialists emerging side by side: An analysis of evolutionary patterns in the cosmopolitan chewing louse genus *Menacanthus*. *International Journal for Parasitology*, *45*, 63–73.

Marussich, W. A., & Machado, C. A. (2007). Host-specificity and coevolution among pollinating and nonpollinating New World fig wasps. *Molecular Ecology*, *16*, 1925–1946.

Mayhew, P. J. (2018). Explaining global insect species richness: Lessons from a decade of macroevolutionary entomology. *Entomologia Experimentalis et Applicata*, *166*, 225–250.

McBride, C. S., Van Velzen, R., & Larsen, T. B. (2009). Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry. *Molecular Ecology*, *18*, 3639–3651.

McKenna, D. D., Sequeira, A. S., Marvaldi, A. E., & Farrell, B. D. (2009). Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 7083–7088.

Medina, R. F., Dickey, A. M., Harrison, K., & Miller, G. L. (2017). Host-associated differentiation in a pecan and water hickory Aphidomorpha community. *Entomologia Experimentalis et Applicata*, *162*, 366–378.

Michell, C., Wutke, S., Aranda, M., & Nyman, T. (2021). Genomes of the willow-galling sawflies *Euura lappo* and *Eupontania aestiva* (Hymenoptera: Tenthredinidae): A resource for research on ecological speciation, adaptation, and gall induction. *G3 Genes|Genomes|Genetics*, *11*, jkab094.

Michell, C. M., Wagner, N., Mutanen, M., & Kyung, M. L. (2022). Dataset for "Genomic evidence for contrasting patterns of host-associated genetic differentiation across shared host-plant species in leaf- and bud-galling sawflies." *Dryad Digital Repository*. https://doi.org/10.5061/dryad.34tmpg4kv

Michell, C. T., & Nyman, T. (2021). Microbiomes of willow-galling sawflies: Effects of host plant, gall type, and phylogeny on community structure and function. *Genome*, *64*, 615–626.

Mlynarek, J. J., & Heard, S. B. (2018). Strong and complex host- and habitat-associated genetic differentiation in an apparently polyphagous leaf mining insect. *Biological Journal of the Linnean Society*, *125*, 885–899.

Murphy, S. M., & Feeny, P. (2006). Chemical facilitation of a naturally occuring host shift by *Papilio machaon* butterfiles (Papilionidae). *Ecological Monographs*, *76*, 399–414.

Nakadai, R., & Kawakita, A. (2016). Phylogenetic test of speciation by host shift in leaf cone moths (Caloptilia) feeding on maples (Acer). *Ecology and Evolution*, *6*, 4958–4970.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274.

Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

Nylin, S., Agosta, S., Bensch, S., Boeger, W. A., Braga, M. P., Brooks, D. R., Forister, M. L., Hambäck, P. A., Hoberg, E. P., Nyman, T., Schäpers, A., Stigall, A. L., Wheat, C. W., Österling, M., & Janz, N. (2018). Embracing colonizations: A new paradigm for species association dynamics. *Trends in Ecology & Evolution*, *33*, 4–14.

Nyman, T. (2002). The willow bud galler *Euura mucronata* Hartig (Hymenoptera: Tenthredinidae): One polyphage or many monophages? *Heredity*, *88*, 288–295.

Nyman, T., Bokma, F., & Kopelke, J.-P. (2007). Reciprocal diversification in a complex plant-herbivore-parasitoid food web. *BMC Biology*, *5*, e49.

Nyman, T., & Julkunen-Tiitto, R. (2005). Chemical variation within and among six northern willow species. *Phytochemistry*, *66*, 2836–2843.

Nyman, T., Leppänen, S. A., Várkonyi, G., Shaw, M. R., Koivisto, R., Barstad, T. E., Vikberg, V., & Roininen, H. (2015). Determinants of parasitoid communities of willow-galling sawflies: Habitat overrides physiology, host plant and space. *Molecular Ecology*, *24*, 5059–5074.

Nyman, T., Onstein, R. E., Silvestro, D., Wutke, S., Taeger, A., Wahlberg, N., Blank, S. M., & Malm, T. (2019). The early wasp plucks the flower: Disparate extant diversity of sawfly superfamilies (Hymenoptera: 'Symphyta') may reflect asynchronous switching to angiosperm hosts. *Biological Journal of the Linnean Society*, *128*, 1–19.

Nyman, T., Vikberg, V., Smith, D. R., & Boevé, J.-L. (2010). How common is ecological speciation in plant-feeding insects? A "Higher" Nematinae perspective. *BMC Evolutionary Biology*, *10*, e266.

Nyman, T., Widmer, A., & Roininen, H. (2000). Evolution of gall morphology and host-plant relationships in willow-feeding sawflies (Hymenoptera: Tenthredinidae). *Evolution*, *54*, 526–533.

Oeyen, J. P., Baa-Puyoulet, P., Benoit, J. B., Beukeboom, L. W., Bornberg-Bauer, E., Buttstedt, A., Calevro, F., Cash, E. I., Chao, H., Charles, H., Chen, M. J. M., Childers, C., Cridge, A. G., Dearden, P., Dinh, H., Doddapaneni, H. V., Dolan, A., Donath, A., Dowling, D., … Niehuis, O. (2020). Sawfly genomes reveal evolutionary acquisitions that fostered the mega-radiation of parasitoid and eusocial Hymenoptera. *Genome Biology and Evolution*, *12*, 1099–1188.

Oksanen, A. J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Simpson, G. L., Solymos, P., Stevens, H., & Wagner, H. H. (2019). *Vegan: Community ecology package*. https://github.com/vegandevs/vegan

Oliveira, J. B. B. S., Faria, M. L., Borges, M. A. Z., Fagundes, M., & de Araújo, W. S. (2020). Comparing the plant–herbivore network topology of different insect guilds in Neotropical savannas. *Ecological Entomology*, *45*, 406–415.

Ord, T. J., & Summers, T. C. (2015). Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evolutionary Biology*, *15*, e137.

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528.

Patten, M. M., Carioscia, S. A., & Linnen, C. R. (2015). Biased introgression of mitochondrial and nuclear genes: A comparison of diploid and haplodiploid systems. *Molecular Ecology*, *24*, 5200–5210.

Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 7495–7500.

Percy, D. M., Argus, G. W., Cronk, Q. C., Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Husband, B. C., Newmaster, S. G., Barrett, S. C. H., & Graham, S. W. (2014). Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? *Molecular Ecology*, *23*, 4737–4756.

Percy, D. M., Page, R. D. M., & Cronk, Q. C. B. (2004). Plant–insect interactions: Double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology*, *53*, 120–127.

Pérez-Pereira, N., Quesada, H., & Caballero, A. (2017). Can parallel ecological speciation be detected with phylogenetic analyses? *Molecular Phylogenetics and Evolution*, *116*, 149–156.

Poveda-Martínez, D., Aguirre, M. B., Logarzo, G., Hight, S. D., Triapitsyn, S., Diaz-Sotero, H., Diniz Vitorino, M., & Hasson, E. (2020). Species complex diversification by host plant use in an herbivorous insect: The source of Puerto Rican cactus mealybug pest and implications for biological control. *Ecology and Evolution*, *10*, 10463–10480.

Poveda-Martínez, D., Varone, L., Fuentes Corona, M., Hight, S., Logarzo, G., & Hasson, E. (2022). Spatial and host related genomic variation in partially sympatric cactophagous moth species. *Molecular Ecology*, *31*, 356–371.

Price, P. W., & Clancy, K. M. (1986). Interactions among three trophic levels: Gall size and parasitoid attack. *Ecology*, *67*, 1593–1600.

Prous, M., Blank, S. M., Goulet, H., Heibo, E., Liston, A., Malm, T., Nyman, T., Schmidt, S., Smith, D., Vårdal, H., Viitasaari, M., Vikberg, V., & Taeger, A. (2014). The genera of Nematinae (Hymenoptera, Tenthredinidae). *Journal of Hymenoptera Research*, *40*, 1–69.

Prous, M., Lee, K. M., & Mutanen, M. (2020). Cross-contamination and strong mitonuclear discordance in Empria sawflies (Hymenoptera, Tenthredinidae) in the light of phylogenomic data. *Molecular Phylogenetics and Evolution*, *143*, e106670.

R Development Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Regnéll, C., Mangerud, J., & Svendsen, J. I. (2019). Tracing the last remnants of the Scandinavian Ice Sheet: Ice-dammed lakes and a catastrophic outburst flood in northern Sweden. *Quaternary Science Reviews*, *221*, 105862.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution*, *19*, 101–109.

Scheffer, S. J., Lewis, M. L., Hébert, J. B., & Jacobsen, F. (2021). Diversity and host plant-use in North American Phytomyza holly leafminers (Diptera: Agromyzidae): Colonization, divergence, and specificity in a host-associated radiation. *Annals of the Entomological Society of America*, *114*, 59–69.

Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593.

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Schmidt, S., Taeger, A., Morinière, J., Liston, A., Blank, S. M., Kramp, K., Kraus, M., Schmidt, O., Heibo, E., Prous, M., Nyman, T., Malm, T., & Stahlhut, J. (2017). Identification of sawflies and horntails (Hymenoptera, 'Symphyta') through DNA barcodes: Successes and caveats. *Molecular Ecology Resources*, *17*, 670–685.

Simmonds, S. E., Fritts-Penniman, A. L., Cheng, S. H., Mahardika, G. N., & Barber, P. H. (2020). Genomic signatures of host-associated divergence and adaptation in a coral-eating snail, *Coralliophila violacea* (Kiener, 1836). *Ecology and Evolution*, *10*, 1817–1837.

Skvortsov, A. K. (1999). *Willows of Russia and adjacent countries: taxonomical and geographical revision*. University of Joensuu.

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

Stireman, J. O., 3rd, Nason, J. D., & Heard, S. B. (2005). Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, *59*, 2573–2587.

Stone, G. N., Hernandez-Lopez, A., Nicholls, J. A., Di Pierro, E., Pujade-Villar, J., Melika, G., & Cook, J. M. (2009). Extreme host plant conservatism during at least 20 million years of host plant pursuit by oak gallwasps. *Evolution*, *63*, 854–869.

Suchan, T., & Alvarez, N. (2015). Fifty years after Ehrlich and Raven, is there support for plant–insect coevolution as a major driver of species diversification? *Entomologia Experimentalis et Applicata*, *157*, 98–112.

Sun, G., Liang, C., Yan, J., Zhang, Y., Shui, Y., & Zhou, Y. (2022). Analysis of complete mitochondrial genome sequence of Pontania dolichura (Hymenoptera: Tenthredinidae) (Thomson, 1871). *Mitochondrial DNA Part B: Resources*, *7*, 905–907.

Sweet, A. D., Boyd, B. M., & Johnson, K. P. (2016). Cophylogenetic patterns are uncorrelated between two lineages of parasites on the same hosts. *Biological Journal of the Linnean Society*, *118*, 813–828.

Tan, A., Abecasis, G. R., & Kang, H. M. (2015). Unified representation of genetic variants. *Bioinformatics*, *31*, 2202–2204.

Vikberg, V., & Zinovjev, A. (2006). On the taxonomy and host plants of North European species of Eupontania (Hymenoptera: Tenthredinidae: Nematinae). *Beiträge zur Entomologie*, *56*, 239–268.

Volf, M., Hrcek, J., Julkunen-Tiitto, R., & Novotny, V. (2015). To each its own: Differential response of specialist and generalist herbivores to plant defence in willows. *The Journal of Animal Ecology*, *84*, 1123–1132.

Volf, M., Pyszko, P., Abe, T., Libra, M., Kotásková, N., Šigut, M., Kumar, R., Kaman, O., Butterill, P. T., Šipoš, J., Abe, H., Fukushima, H., Drozd, P., Kamata, N., Murakami, M., & Novotny, V. (2017). Phylogenetic composition of host plant communities drives plant-herbivore food web structure. *The Journal of Animal Ecology*, *86*, 556–565.

Wachi, N., Matsubayashi, K. W., & Maeto, K. (2018). Application of next-generation sequencing to the study of non-model insects. *Entomological Science*, *21*, 3–11.

Wagner, N. D., Gramlich, S., & Hörandl, E. (2018). RAD sequencing resolved phylogenetic relationships in European shrub willows (*Salix* L. subg. *Chamaetia* and subg. *Vetrix*) and revealed multiple evolution of dwarf shrubs. *Ecology and Evolution*, *8*, 8243–8255.

Wagner, N. D., He, L., & Hörandl, E. (2020). Phylogenomic relationships and evolution of polyploid *Salix* species revealed by RAD sequencing data. *Frontiers in Plant Science*, *11*, e1077.

Ward, A. K. G., Bagley, R. K., Egan, S. P., Hood, G. R., Ott, J. R., Prior, K. M., Sheikh, S. I., Weinersmith, K. L., Zhang, L., Zhang, Y. M., & Forbes, A. A. (2022). Speciation in Nearctic oak gall wasps is frequently correlated with changes in host plant, host organ, or both. *Evolution*, *76*, 1849–1867.

Waters, J. M., & McCulloch, G. A. (2021). Reinventing the wheel? Reassessing the roles of gene flow, sorting and convergence in repeated evolution. *Molecular Ecology*, *30*, 4162–4172.

Weber, M. G., Wagner, C. E., Best, R. J., Harmon, L. J., & Matthews, B. (2017). Evolution in a community context: On integrating ecological interactions and macroevolution. *Trends in Ecology & Evolution*, *32*, 291–304.

Weinstein, S. B., & Kuris, A. M. (2016). Independent origins of parasitism in Animalia. *Biology Letters*, *12*, 20160324.

Windsor, D. A. (1998). Most of the species on Earth are parasites. *International Journal for Parasitology*, *28*, 1939–1941.

Wu, J., Nyman, T., Wang, D.-C., Argus, G. W., Yang, Y.-P., & Chen, J.-H. (2015). Phylogeny of *Salix* subgenus *Salix* s.l. (Salicaceae): Delimitation, biogeography, and reticulate evolution. *BMC Evolutionary Biology*, *15*, e31.

Yamaguchi, H., Tanaka, H., Hasegawa, M., Tokuda, M., Asami, T., & Suzuki, Y. (2012). Phytohormones and willow gall induction by a gall-inducing sawfly. *The New Phytologist*, *196*, 586–595.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Michell, C. T., Wagner, N., Mutanen, M., Lee, K. M., & Nyman, T. (2023). Genomic evidence for contrasting patterns of host-associated genetic differentiation across shared host-plant species in leaf- and bud-galling sawflies. *Molecular Ecology*, *00*, 1–19. https://doi.org/10.1111/mec.16844