Preprint (2025), **1**, 1–27

REVIEW TYPE

Towards a quantitative view of the NLR gene family evolution in the genome space

Luzie U. Wingen,[*†] Duncan Crosbie,[‡] Yiheng Hu,[¶] Eric Kemen,[¶] Xinyi Liu,[§] Marion C. Müller,[§] Niklas Schandry,[‡] Korbinian Schneeberger,[|0] Detlef Weigel,[⊥] and Aurélien Tellier[*†]

†Professorship for Population Genetics, School of Life Sciences, Technical University of Munich, 95354 Freising, Germany

‡Faculty of Biology, Ludwig-Maximilians-Universität München, 82152 Martinsried, Germany

¶Centre of Plant Molecular Biology (ZMBP), Eberhard-Karls-University of Tübingen, 72076 Tübingen, Germany

§Chair of Phytopathology, School of Life Sciences, Technical University of Munich, 95354 Freising, Germany

|0Institute for Crop Biology, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany; (2) Max Planck Institute for Plant Breeding Research, Cologne, 50829 Cologne, Germany; (3) Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine University, Düsseldorf, Germany.

⊥Department of Molecular Biology, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany

*Corresponding author. Email: luzie.wingen@tum.de; aurelien.tellier@tum.de

Abstract

Plants and their pathogens coevolve over long time periods, and the history of coevolution is recorded in plant genes that confer pathogen resistance. Many of these code for Nucleotide-binding Leucine rich Repeat proteins (NLRs), which are crucial for distinguishing friends from foes and triggering potent defense responses. Advances in the ability to sequence genomes from many different species as well as many genomes from the same species reveal that 1) the number of NLR genes differ widely between species, and 2) NLR genes may exhibit extensive nucleotide variation as well as presence/absence polymorphism within species. The "birth-and-death process" is thus a generally useful framework for describing NLR gene evolution. In the light of the latest insights into the genomic features associated with NLR gene diversity, we aim here to evaluate the contribution of forces involved in NLR gene family evolution and presence/absence variation: mutation, recombination, gene duplication and deletion, and natural selection. To do so, we highlight novel combinations of population genomics methods and statistics that can provide an improved framework for describing NLR gene family evolution in the genome space, especially accounting for selective processes stemming from their function within resistance gene networks and the resulting quantitative defense response.

Keywords: population genomics, selection, neutral evolution, resistance, susceptibility

Introduction

With advances in sequencing technologies, especially single molecule long reads with rapidly declining error rates, we are gaining an unprecedented view of the organization of plant genomes and gene families. This resolution of genome assemblies generated with long reads not only surpasses those generated with short reads but also those produced with legacy technologies such as BAC tiling paths and Sanger sequencing. Such assemblies reliably and accurately distinguish individual gene copies of multi-copy gene families, resolving the diversity between individual copies, while also uncovering the extent of large and small genomic arrangements. This allows one to summarize the occurrence of gene copy number variation (CNV) and presence-absence variation (PAV) within and between populations and paves the way for pan-genomic analyses at the intra- and inter-specific levels (Table 1). However, a comprehensive understanding of the evolutionary mechanisms driving gene families in the genome space is still lacking, as theory and methods of analysis currently lag behind the acquisition of long-read sequencing data. Indeed, while population genetics theory has been very successful in developing statistical methods, for example, using diffusion and coalescence models, to explain patterns of polymorphisms in unique regions, it largely fails when it comes to proposing mechanistic models for how more complex regions of the genome diversify. A major reason for this is the reliance on the Markovian assumption of recombination to model genealogies along the genome, which is not fully accounting for additional processes including duplications, inversions, deletions, conversions, non-equal recombination (that is intergenic recombination between paralogs), and transpositions. Therefore, we argue that new statistical methods, firmly grounded in an extended theory of evolutionary and population genomics, are required. Those methods would aim to infer and dissect the importance of the various forces shaping gene family evolution, explaining the dynamics of genomic change within and across species. To motivate such developments, we focus here on a case study of the NLR (NBS-LRRs, Nucleotide-binding Leucine rich Repeat proteins) gene family in plants, which is highly variable in gene copy number and well-studied within and across-taxa.

NLRs are an important component of the plant immune system (Pradeu et al. 2024). Like other organisms, plants interact with a diverse array of microbes and invertebrates, entering into mutualistic, commensalistic or parasitic relationships. Of particular interest are repeated cycles of antagonistic interactions with pathogens, as encapsulated by the “Red Queen” metaphor (Van Valen 1973, Dawkins and Krebs 1979). The resulting reciprocal adaptation, or coevolution (Thompson 2005, Burdon and Laine 2019, Milgroom 2023), occurs on the host side at genes responsible for pathogen recognition and defense activation, including the NLR genes, and on the pathogen side in genes for so-called effectors, which are recognized by NLR proteins, as well as other genes

responsible for pathogen success (Dodds and Rathjen 2010). To understand the factors that play a role, it is essential to distinguish between the different timescales at which processes occur (Table 1). At short timescales of up to a few years, epidemiological dynamics (and environmental conditions) define disease prevalence and severity, and thus selective pressures (Agrios 2024, Milgroom 2023, Burdon and Laine 2019). This timescale is the realm of so-called boom-and-bust cycles of pathogen populations observed in agriculture (Agrios 2024), and of ecological-evolutionary (eco-evo) dynamics in wild systems (Boots et al. 2014). These eco-evo cycles generate neutral and selective evolutionary forces that act over a period of a few years to a few millennia (Tellier, Moreno-Gámez, and Stephan 2014, Živković et al. 2019). The resulting genomic footprints can be studied within the framework of conventional population genetics, applying well-established theoretical concepts to within-species polymorphism data. Extending this timescale up to millions of years yields the evolution of genome structure and genome space, which can be observed by following between-species divergence and synteny (Lynch 2007). The birth-and-death scenario proposed by Michelmore and Meyers over two decades ago (Michelmore and Meyers 1998) provides a general coevolutionary framework for understanding the evolution of NLR gene families and individual NLR gene copies, linking these different ecological, evolutionary and genomic timescales. The fast evolution and high turnover of NLR genes is reflected in an above genome average nucleotide diversity (e.g. Bakker et al. 2006, Clark et al. 2007, Prigozhin and Krasileva 2021, Stam et al. 2019), varying numbers of partial or complete gene copies (copy number variation, CNV) and a high level of presence-absence variation (PAV) (Lee and Chae 2020,Van de Weyer et al. 2019, Teasdale et al. 2025,Silva-Arias et al. 2025).

In the following, we first describe the classic evolutionary and genetic mechanisms that underpin and fuel the NLR gene birth-and-death process (Michelmore and Meyers 1998): mutation, recombination, genetic drift, selection, and gene duplication/deletion. Second, we evaluate the importance of these processes in the light of recent findings made available by improved genome data quality in plants (especially in *Arabidopsis thaliana*). Concomitantly, we suggest a set of statistical analyses and guidelines for studying NLR gene diversity (at the homolog and nucleotide levels) and inferring the relative importance of the various acting evolutionary forces. We also highlight open questions arising from these new studies. Finally, based on recent theoretical work and new insights on the molecular characterization of NLR gene networks, we propose an update to the definitions and effects of population genetics selective processes. We therefore hope to show a way for how to develop an integrated eco-evo quantitative view of NLR (and other) gene family evolution in the plant genome space.

What is the NLR birth-and-death process?

The Red Queen hypothesis posits that rapidly evolving pathogens that make use of multiple, redundant and overlapping mechanisms to colonize their hosts lead to continuous cycles of reciprocal adaptation, without either the host or the pathogen prevailing in the long term. As the host evolves resistance, the pathogen jettisons genes for molecules recognized by the host, deploying instead alternative molecules for successful infection. In response, the host evolves new means that can suppress the pathogen until the pathogen again changes its genetic makeup. The defining feature is negative frequency-dependent selection generating allele frequencies that fluctuate over time. In the Red Queen scenario (or trench warfare) genetic diversity is maintained in both host and pathogen. In contrast, under the arms race situation, both host and pathogens rely on continuous innovation rather than recycling of existing variation. While the timescales over which alleles persists in a population differs under the two scenarios (Tellier, Moreno-Gámez, and Stephan 2014,Brown and Tellier 2011), common to both situations is that new genetic diversity needs to be generated. This variability can be generated by various mechanisms: 1) Mutation at the nucleotide level (point mutation, small insertion-deletions). 2) Recombination reshuffling variability between homologous copies of NLRs, with unequal recombination reshuffling variation between recently diverged copies in the same gene cluster, and gene conversion doing the same, but between more distant paralogous copies. 3) Gene duplication or deletion occurring by DNA break and repair, sometimes associated with the activity of transposable elements (TEs), generating presence/absence variation (PAV) at single genes or duplicated copies. 4) Unequal recombination between sets of paralogous copies further enhancing copy number variation (CNV) and driving the expansion of the gene family. Variation is removed primarily by genetic drift but also by costs such as autoimmunity, which can be the consequence of NLR gene truncations, NLR overexpression or inappropriate interactions between NLR and other immune genes (Freh et al. 2021).

Three types of selection occur at a given NLR locus. Positive selection, due to negative indirect frequency-dependent selection (FDS; Tellier and Brown 2007), generates selective sweeps when an NLR variant provides an advantage, for example, by activating effective defense against a common pathogen. FDS is indirect, as the fitness of host alleles depends on the parasite allele frequencies and *vice and versa*, and negative, as rare alleles have an advantage against common alleles. This yields over several cycles of selection the so-called arms race dynamics (Bergelson et al. 2001, Woolhouse et al. 2002). Negative selection (or purifying selection) removes deleterious variants, for example, if mutations at a given NLR locus incur a cost of resistance (Tian et al. 2002, Freh et al. 2021) or activate auto-immunity (Freh et al. 2021). Balancing selection, due to stable coevolutionary cycles

(fluctuating selection) driven by negative indirect FDS in large populations, can generate the so-called trench warfare dynamics at the NLR locus (Stahl et al. 1999). While costs of resistance and virulence are necessary for balancing selection to occur, these are not sufficient and additional characteristics of natural systems (negative direct FDS, Tellier and Brown 2007) are required such as spatial and temporal heterogeneous selection (review in Brown and Tellier 2011), diffuse coevolution (Karasov et al. 2014) or epidemiological feedbacks (Boots et al. 2014). Negative direct FDS occurs when the fitness of an allele depends negatively on its own frequency (Tellier and Brown 2007).

These evolutionary processes and mechanisms drive the birth-and-death process of NLR genes, thereby shaping variation in the composition of the NLR gene family between individuals, populations and species. When the birth of an NLR occurs via gene duplication, the fate of each paralogous copy depends on these neutral and selective processes (e.g. Innan and Kondrashov 2010). Balancing selection at one NLR locus for several resistance alleles can favour gene duplication and the subsequent subfunctionalization of each copy to recognize different pathogen effectors or effector alleles (Van der Hoorn, De Wit, and Joosten 2002). Neo-subfunctionalization can also occur through selection at one copy for recognition of new pathogen variants (arising by mutation, recombination or gene conversion). In both cases, diversifying selection occurs and is observed across the different NLR paralogous gene copies (Innan and Kondrashov 2010, Otto, Zheng, and Wiehe 2022). Pseudogenization (by mutation, recombination or gene conversion) may inactivate one NLR copy, which becomes free to evolve a new function or may be lost. Ultimately, an NLR copy undergoes gene death either by deletion or by non-functional pseudogenization following genetic drift or selection against deleterious copies/alleles (Michelmore and Meyers 1998). Copies/alleles become deleterious when they exhibit high fitness costs that are not compensated by providing *ad hoc* resistance to a pathogen (Stahl et al. 1999,Tellier and Brown 2007), or when they inadvertently trigger autoimmune responses (Freh et al. 2021). In this framework, the number of NLR gene copies in the genome and their polymorphism (at the CNV levels, *a.k.a.* the pan-NLRome) depends thus on the balance between the rates of gene birth and death. For the plant host to cope with the fast adaptation of their pathogens, it is classically expected that higher mutation, recombination, and gene duplication rates may be advantageous in generating more variable NLR genes, which in turn would promote higher levels of CNV reflecting the NLR gene turnover rate.

How to identify structural NLR variation in the genome?

The first task to study NLR diversity and dynamics is to identify NLR variation. While whole-genome short read sequencing or NLR Ren-seq coupled with short or long reads provides some

insights into NLR CNVs (e.g. Stam et al. 2019, Van de Weyer et al. 2019, Seong et al. 2020, Silva-Arias et al. 2025, S.-X. Li et al. 2025), such analyses suffer from mapping uncertainty and reference biases. The clustering of NLR genes in the genome and the quantification of the family members and the nucleotide diversity can be assessed based on their NB-ARC domain (Prigozhin and Krasileva 2021, Silva-Arias et al. 2025); but the resolution of CNV remains error-prone (Silva-Arias et al. 2025). Long-read assemblies are much better at disentangling paralogous sequences, with the caveat that near-identical copies, especially in heterozygous and polyploid genomes, continue to be challenging (Cheng et al. 2021, Rhie et al. 2021, Rautiainen et al. 2023). The use of new bioinformatic methods for pan-genomes including massive pair-wise comparisons (Igolkina et al. 2025) and genome graphs (Garrison et al. 2024) has become instrumental in accounting for genome rearrangements occurring in the NLR gene family (see the recent review by Bao and Weigel 2025). However, while genome graphs are useful representations, they describe only the outcome of evolution and do not allow directly for the inference of the causal mechanisms and evolutionary forces underpinning a given graph. We suggest that an important aim of future research must be to integrate genome graphs with population genomics theory – which should also improve parameter choices that have to be made when building genome graphs.

An example of how long-read assemblies improve our understanding of the genomic landscape of NLR genes comes from a comparison of 17 *A. thaliana* accessions (Teasdale et al. 2025). This study introduced the concept of an NLR neighborhood, defined as a region that contains at least one NLR in at least one of accession. It revealed a higher-than-expected density of transposable elements (TEs), which may indicate a role of TEs in gene duplication and/or deletion. So far, this is, however, merely a correlation: are NLR loci more prone to attracting TE insertions or allowing for TE mobilization, or is TE fixation driven by hitch-hiking (see below linked selection) due to evolutionary (selective) changes at NLR genes? Furthermore, NLR gene families appear varied in their diversity (Teasdale et al. 2025), and there do not appear to be single metrics that fully capture NLR diversity. For example, the *RPP13* gene is highly diverse in sequence but not in isoforms or copy number, whereas the *DAR4* gene exhibits the opposite pattern, while the *RPP4* gene consistently shows extreme values for most metrics. This suggests that certain NLR genes and families have perhaps primary roles in dealing with slowly changing pathogen threats, while others have higher potential for adaptability in the face of fast evolving pathogens (Bakker et al. 2006, Laflamme et al. 2020, Prigozhin and Krasileva 2021, Sutherland et al. 2024 in *A. thaliana*, Stam et al. 2019 in *Solanum chilense*). A central aim of the present study is to provide a roadmap for revealing and quantifying this heterogeneity in NLR gene evolutionary trajectories and the underlying mechanisms.

Is the high genetic diversity of NLR genes due to higher mutation rate?

Based on phylogenetic reconstruction of *A. thaliana* NLR families from 62 long-read assemblies of NLR loci, Prigozhin and Krasileva 2021, define two categories of NLR genes: 1) highly variable NLR genes (hvNLRs) encoding proteins with unusually high amino acid diversity and 2) low variability NLR genes (non-hvNLRs), which presumably experience stronger purifying selection. A follow-up-study (Sutherland et al. 2024) found that hvNLRs exhibit higher expression levels, lower gene body cytosine methylation, and closer proximity to TEs than non-hvNLRs. As expected, hvNLRs have elevated synonymous and nonsynonymous nucleotide diversity (*i.e.* πS and πN, respectively, see Table), higher Tajima's D values (see Table), and are preferentially found in chromatin states associated with an increased probability of per-site mutation (based on the estimates in Monroe et al. Nature 2022). While purifying (negative) selection to maintain a specific function is likely weaker at hvNLRs than non-hvNLRs (along with the associated background selection, see below), these analyses do not tell us whether balancing selection or positive selection drives hvNLR diversity, as relaxed purifying selection (and weak linked selection) or neutral evolution would generate similar footprints of nucleotide diversity. An analysis of the paralogous assignments and clustering of hvNLRs is required (for which genome graphs could help, see above) to assess the duplication (paralogous) status of the hvNLRs, as duplicated genes should more easily accumulate mutations due to (neo-)subfunctionalization (Innan and Kondrashov 2010). In addition, as more complete genomes become available, it will become possible to ask whether hvNLR loci feature individual alleles (haplotypes) that undergo purifying selection, i.e., whether diversity (and the speed of evolution) at these loci is discrete or continuously distributed. We also suggest that quantitative analyses of the selection strength can now be performed both for entire NLR subfamilies (e.g., TIR-NLRs, CC-NLRs, CCr-NLRs) or categories (hvNLRS versus non-hvNLRs, single genes or clustered genes, or canonical NLRs and truncated NLRs) via the estimation of the distribution of fitness effects (DFE, Keightley and Eyre-Walker 2007) and the percentage of adaptive mutations (method in Sendrowski and Bataillon 2024, applications in Grandaubert, Dutheil, and Stukenbrock 2019). While DNA mutations are generally considered to be non-targeted with respect to their genomic location and their functional consequence, there are rare cases in which targeted rearrangements of the DNA sequence occur, which enhance the probability for adaptive changes (review in Hanlon, Cagan, and Eves-van den Akker 2025). Nonetheless, based on published results (Van de Weyer et al. 2019, Prigozhin and Krasileva 2021, Sutherland et al. 2024), we cannot yet conclude whether or not the latter applies to NLR genes. We thus caution against the general claim that NLR genes exhibit higher per-site mutation rates than the rest of the genome.

Are meiotic recombination and gene conversion specifically elevated at NLR loci?

Following the tenets of the Red Queen hypothesis (Van Valen 1973, Barton and Charlesworth 1998), sexual reproduction, and especially meiotic recombination (introducing chromosome arm exchanges and gene conversions), is expected to be an important evolutionary mechanism for generating new genotypes in hosts (and pathogens). The shuffling of existing nucleotide diversity generates novel multi-locus haplotypes with every generation. Moreover, if recombination occurs within clusters or genes, it can even create new NLR cluster or allelic variants. The situation for NLRs remains, however, unclear, as recent insights challenge the idea that increased meiotic recombination is underlying the high diversity of NLR genes (or gene clusters) and thereby drive accelerated evolution of NLR genes. Studies in *A. thaliana* have shown that the frequency of meiotic crossovers in R-gene regions does not deviate greatly from genome-wide average (Choi et al. 2016). Although meiotic crossovers occur in NLR gene clusters, occasionally even forming recombination hotspots, the overall crossover rates within these clusters typically align with genome-wide averages (Choi et al. 2016). Notably, structural rearrangements, which are hallmarks of complex NLR gene clusters (Jiao et al. 2021), inhibit recombination, resulting in extensive recombination suppression across these regions. The lack of elevated crossover frequency, coupled with the likely suppression of crossovers due to massive structural variation, suggests that crossovers recombination alone may not be the main driver behind the rapid diversification of NLR genes in *A. thaliana*. Alternative mechanisms involving gene duplication and copy number variation as prerequisites, such as (meiotic or somatic) gene conversion (Hörger et al. 2012 in wild tomato) between NLR paralogs or unequal recombination (e.g., as seen at MHC loci in animals, Otto, Zheng, and Wiehe 2022), may be additional drivers of NLR gene evolution. Yet, their occurrence and frequency have to be quantified with high-quality genome assemblies, preferably using a collection of genomes from both more closely and more distantly related accessions, since excessive structural variation will reduce opportunities for unequal crossovers. Note that population genetics models of gene conversion (Innan 2003, Thornton 2007) and unequal recombination (Otto, Zheng, and Wiehe 2022) exist and can be used to generate predicted polymorphism patterns (SNPs) across paralogs, which in turn can be used to infer the rate of gene conversion (Hörger et al. 2012, Table).

How many different types of selection act at NLR genes?

Since the discovery of plant resistance loci encoded by NLR genes, numerous studies have utilized polymorphism data to evaluate the occurrence of positive selection (arms race) versus balancing selection (trench warfare dynamics) both at functionally defined loci as well as NLRs genome wide.

To keep the list of references to a reasonable number, we mention here only a few systems and studies. Early efforts based on PCR amplification and Sanger sequencing have been conducted in *A. thaliana* (Stahl et al. 1999, Bergelson et al. 2001, Bakker et al. 2006), wild tomato species (Rose, Michelmore, and Langley 2007, Hörger et al. 2012) or common bean (De Meaux et al. 2003). These initial studies found clear footprints of selection only at a few loci. With the advent of new sequencing technologies and improved methods for carrying out selection scans (see Table), this picture has not significantly changed (see Q. Wu et al. 2017, Stam et al. 2019, Wei, Silva-Arias, and Tellier 2023). As predicted by theory (Brown and Tellier 2011, Tellier, Moreno-Gámez, and Stephan 2014), arms races may not be that common in natural systems, which is very different from agricultural crop systems. Looking at this from the other side, it was found that pathogen population of wild plants might indeed be stable over very long time scales, with diversity among them driven as much by pathogen-pathogen antagonism as by plant-pathogen interactions (Shalev et al. 2022, Backman et al. 2024). However, typical polymorphism footprints of balancing selection are more subtle to uncover (than selective sweeps), as they depend on the local recombination rate and the time during which selection acted to maintain variation (Stahl et al. 1999,Tellier, Moreno-Gámez, and Stephan 2014, Charlesworth and Jensen 2021). In other words, balancing selection is more easily uncovered if it underlies variation that is maintained across species after speciation. On the flip side, such signals can be difficult to detect, because recombination over a long time will reduce the size of the segments generating the signals; in extreme cases, only a single nucleotide polymorphism might be left, which can be difficult to distinguish from technical artifacts andrecurrent mutations. A special situation was described for an outcrossing species, *Capsella grandiflora*, and a sister species, *C. rubella*, that had only very recently diverged from it, such that shared blocks of polymorphism were still large (Gos, Slotte, and Wright 2012, Bachmann et al. 2019, Koenig et al. 2019). Trans-species SNPs are very unevenly distributed along the genome and strongly enriched in immune-related loci, strongly suggesting balancing selection at NLR and other immune genes. Many of the trans-species SNPs could be reproduced in a comparison with a third, also selfing species, *C. orientalis*, that had apparently undergone a very recent, very strong genetic bottleneck (Koenig et al. 2019). While intriguing, a note of caution comes from a study that showed how the heterogeneous distribution of genomic regions with high nucleotide diversity (and ancestral polymorphism) in the genome of selfing species could also be attributed to the process of transition from outcrossing to selfing (Strütt et al. 2023). This means that such footprints of trans-species polymorphism may also be generated by other processes, such as a transition in reproductive mode, without the action of balancing selection. The original analysis (Koenig et al. 2019) was based on short-read data, and reconstructing entire haplotypes with long

reads should greatly help to further resolve the mutational history and structural variation at the loci in question and confirm what type of selection has acted on these.

What is the effect of linked selection around NLR genes?

A consequence of the heterogeneity in recombination rates (Choi et al. 2016, Jiao et al. 2021) and in the strength of positive, negative and balancing selection along the genome is that the genomes is replete with regions that experience linked selection due to hitch-hiking and background selection, which in turn affects local nucleotide diversity and population genetics statistics beyond the focal loci that are under selection (Charlesworth and Jensen 2021, Table). The effect of linked selection is more pronounced in species with long-range linkage disequilibrium, such as selfing species (*A. thaliana*, *Arabis* ssp., *Capsella rubella*), compared to outcrossers (*Solanum chilense*, *Capsella grandiflora*). Linked selection has several, often overlooked consequences for the study of NLR genes, gene clusters and duplicates.

First, linked selection leads to per-site recombination rate and nucleotide diversity becoming correlated (demonstrated in Begun and Aquadro 1993 and in *A. thaliana* by Lian et al. 2022). This correlation stems from meiotic recombination reducing the effect of linked selection (decreased linkage disequilibrium) around the selected sites (Begun and Aquadro 1992, Charlesworth and Jensen 2021). The effect (Begun and Aquadro 1993, Charlesworth and Jensen 2021). The effect of linked selection can be, for example, uncovered by analyzing measures of synonymous to non-synonymous polymorphisms ($\pi N/\pi S$, Table) while accounting for local genomic recombination rates (Campos et al. 2014). When recombination is low, higher linkage disequilibrium and stronger linked selection generate so-called genetic draft around selected loci by which a genomic region exhibits footprints of neutral evolution (Table), albeit with reduced nucleotide diversity (Achaz and Schertzer 2023). Furthermore, reduced recombination enhances the Hill-Robertson effect (Hill and Robertson 2007) by which selection at the focal locus influences the efficacy of selection at neighboring loci/genes. In other words, a locus under positive selection or negative selection can influence adjacent loci that are under the opposite type of selection. The polymorphism pattern in the adjacent loci would thus not produce the expected footprints of selection (Table). Deciphering the evolutionary history of NLR genes, which can be found in tandem arrangements or in gene clusters, may prove to be particularly difficult in species with long-range LD. The correlation in nucleotide diversity or other neutrality tests (Table) between tandem or duplicated NLRs should be interpreted with caution, as it may not reflect co-adaptation of genes but rather correlation due to linked selection. When analyzing NLR gene evolution over phylogenetic timescales, recombination and linked selection also play a

role, affecting the accuracy of phylogenetic reconstruction of the history of tandem NLR genes or members of NLR gene clusters. We note that the sensitivity of phylogenetic methods to infer selection via the dN/dS (Table) ratio does depend on recombination rate (Kryazhimskiy and Plotkin 2008). To our knowledge, it has not yet been fully assessed how sensitive to recombination (and potential biases) the tools are that are used to infer episodic measures of selection along a phylogeny (as used in Sutherland et al. 2024). Accounting for recombination rate (and linked selection) is thus a prerequisite for studying the possible co-adaptation of NLR genes in tandem or in clusters. For example, the strength, prevalence, and consequences of linked selection and co-adaptation between NLR genes are likely to differ between *A. thaliana*, a selfing species, and the outcrossing wild tomato *S. chilense* (see discussion in Silva-Arias et al. 2025).

How many NLR genes are pseudogenized?

A key feature of the birth-and-death process is the intermediate step of pseudogenization of NLR genes. The assessment of pseudogenization faces, however, several issues. First, there are many examples of both recently truncated NLR alleles or of *bona fide* NLR genes that lack many, or even most of the canonical NLR domains, such as TIR-only genes, that are clearly functional (Barragan and Weigel 2021). In addition, while the progressive accumulation of mutations might lead to eventual complete removal of a gene from the genome, pseudogenized NLR genes could act as a reservoir of diversity (for subsequent neofunctionalization). It has been investigated whether pseudogenized NLR genes are likely to evolve neutrally by comparing the ratio of non-synonymous to synonymous changes in genes and pseudogenes in *A. thaliana* and cultivated rice (Zou et al. 2009). While not all NLR genes could be resolved in this study due to the lack of long-read sequencing data, the authors observed 1) that many genes annotated as pseudogenes are actually subject to purifying selection, 2) that some apparent pseudogenes are expressed (albeit at a low level), and 3) that there is a positive correlation between the size of gene families with different functional domains and the number of pseudogenes. We suggest that by combining the methods described above and in Table , we can dissect the various evolutionary forces driving evolution of NLR genes that may or may not be true pseudogenes and infer the rate of neofunctionalization.

What is the relative importance of neutral versus selective forces?

The efficacy of any type of selection depends on the effective population size, that is, the amount of genetic drift. This is encapsulated in the drift barrier concept (Lynch 2007), which states that the evolution of. mutation rates is constrained by the opposing effects of natural selection and genetic drift.

High mutation rates are deleterious to the genome, and low mutation rates decrease a population's adaptability (Lynch 2007, Hanlon, Cagan, and Eves-van den Akker 2025). Therefore, it can be expected that an optimal mutation rate is reached for the genome of a given species, and possibly for different types of genes, if genetic drift is not too strong. In *A. thaliana*, the mutation rate may be only marginally higher for NLR genes than for other genes in the genome, if at all (see above), but this rate remains unknown for other plant species. We speculate that the small local population size and high selfing rate in *A. thaliana* generate relatively strong genetic drift (and LD), possibly limiting the ability of fine tuned mutation rates at NLR genes to evolve. We also note that the efficacy of selection for positive mutations and against deleterious mutations (purifying selection) is likely to be weaker in *A. thaliana* compared to outcrossing plant species. This could mean that mildly deleterious (costly) NLR gene copies may remain in the genome for extended periods, appearing thus as segregating PAVs/CNVs in *A. thaliana*, whereas they would be more quickly lost from the entire population in outcrossers with stronger purifying selection. In addition, advantageous NLR genes that confer pathogen resistance may not always become rapidly fixed (and remain polymorphic as CNVs) for a long period if local selection is weak compared to drift. In addition, certain combinations of alleles at unlinked NLRs can be deleterious (Bomblies and Weigel 2007), which could increase the strength of purifying selection against NLRs that have otherwise no positive function in the organism. In species that mostly self, this effect is expected to be negligible, but it might be more relevant in obligate outcrossers.

A second process to consider is meiotic recombination. As for mutation, there are advantages and disadvantages to recombination (Barton and Charlesworth 1998), and the optimum rate of homologous recombination at NLR genes must be determined by the strength of genetic drift (as it constrains the efficacy of selection). However, contrary to the Red Queen hypothesis, meiotic recombination in *A. thaliana* does not appear to be higher at NLR genes than in the rest of the genome (Choi et al. 2016), and thus other mechanisms such as gene duplication, deletion, and unequal recombination may be necessary to generate the observed NLR gene clusters and large NLR gene families. NLR gene duplicates need to diverge by mutation to undergo subfunctionalization or neosubfunctionalization (see above), a process that is slowed down by gene conversion and by the low effective population size and thus mutational input in *A. thaliana*. These various mechanisms (Figure 1) collectively provide a general explanation for the large number of CNVs found at NLR genes in *A. thaliana* (Van de Weyer et al. 2019, Lee and Chae 2020, Teasdale et al. 2025) versus outcrossing species (such as *S. chilense*; Silva-Arias et al. 2025).

How do NLR genes evolve in the context of gene networks and clusters?

Ultimately, we aim to integrate the previous insights from population genomics with recent molecular insightss on the function of NLR genes (Figure 1). As NLR genes function as part of gene networks, one can predict heterogeneity in selective constraints across genes. For example, more conserved NLR genes (such as the genes coding for the NRC clade; C.-H. Wu et al. 2017; Stam et al. 2019; Sakai et al. 2024) are likely to be more pleiotropic and have more connections to downstream signalling (and be under stronger purifying selection), whereas sensor NLR genes involved in effector recognition may be more peripheral in such networks (and under weaker purifying selection). NLR gene network' structures are likely organized around host genes coding for key guardee proteins that are targeted by pathogen effectors for their function in plant physiology, development or defense response. The type of genes involved in defense response such as those coding for guardee (an effector target), guard (resistance protein monitoring a guardee) and decoy (resistance protein which chiefly recognizes pathogen effector) could thus be understood in the context of gene networks, multilocus epistatic interactions and trade-offs for effector binding and signaling (Weiner et al. PNAS 2025). This epistasis between NLR genes defines the strength and type of selection occurring at these loci (positive, negative, and balancing) as well as the inherent evolutionary constraints. We speculate that neofunctionalization of gene copies of more connected NLR genes in the networks may explain the evolution of guard and decoy coding genes, as gene duplication can free one copy for further specialization and refinement of the effector detection function (Van der Hoorn, De Wit, and Joosten 2002, Dodds and Rathjen 2010, Adachi, Derevnina, and Kamoun 2019). This process is arguably most visible in the evolution of Integrated Domains (IDs), where duplicated guardee domains are integrated directly into the NLR architecture (Yang et al. 2022, Y. Li et al. 2024). Duplicated gene copies of peripheral genes of the NLR networks may also (neo-)subfunctionalize if selection pressure is strong enough. Duplicated pairs or paralog NLRs, which subfunctionalize, could be found as genes coding for sensor and helper. In other words, the topology of these functional networks would determine the physical arrangement of NLRs in the genome, for example co-functioning pairs (sensor/helper) would be maintained in tight physical clusters (head-to-head orientation). Due to the relatively high linkage disequilibrium and linked selection in *A. thaliana*, the sensor and helper genes are expected to exhibit similar population genetics statistics (Table , Van de Weyer et al. 2019). In line with this idea, there is evidence for the evolution of NLR gene clusters and duplicates starting from a duplicated pair of sensor and helper NRCs in the Asteroid family (Sakai et al. 2024).

Following up from the organization of NLR gene networks (C.-H. Wu et al. 2017, Sakai et al. 2024), the quantitative nature of NLR recognition (Weiner et al. 2025) and the evolutionary

theory of gene network evolution (Wagner 1996, Fierst and Phillips 2015 and reference therein, Wei et al. 2024, Pouzet and Le Rouzic 2024), we suggest two hypotheses that can be tested using evolutionary analyses of gene families. 1) The change of core (well-connected and pleiotropic) NLR genes (such as NRC genes) in networks would allow large steps of adaptation, for example recognition of new pathogens with new host target gene (guardee), but at the price of possible loss of signaling competence for upstream sensors and/or of control of the resistance pathway (yielding autoimmunity). 2) The change of peripheral (less connected) NLR genes would result in recognizing additional pathogen effectors without destabilizing the networks, but these changes would be limited to altering recognition specificity (e.g., via LRR variation or ID acquisition). NLR gene duplication, mutation, recombination, and CNV would then be under the selective constraint defined by the structure and robustness of the NLR defense networks and the quantitative nature of NLR recognition and activation (Weiner et al. 2025).

Conclusion

Our contribution aims to pave the way for a theory of the evolution of genomic space at intra- and inter-population and species levels, with specific application to gene families (NLR genes) in plants. We emphasize the need for analyses of high-quality (long-read assembly) datasets, linking intra-population to inter-species pan-genomes, building on new bioinformatic tools and emerging topics in population genomics theory.

Acknowledgenments

All authors acknowledge support by the TRR356/I "Genetic Diversity Shaping Biotic Interactions of Plants" (491090170) of the German Research Council (DFG). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability statement

There is no data affiliated with this manuscript.

References

Achaz, G, and E Schertzer. 2023. Weak genetic draft and Lewontin's paradox. *bioRxiv,* https://doi.org/10.1101/2023.07.19.549703.

Adachi, Hiroaki, Lida Derevnina, and Sophien Kamoun. 2019. NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Current Opinion in Plant Biology* 50:121–131. https://doi.org/10.1016/j.pbi.2019.04.007.

Agrios, GN. 2024. *Plant pathology.* Edited by Richard Oliver. New York: Academic Press Inc. ISBN: 978-0-12-822429-8.

Bachmann, Jörg A., Andrew Tedder, Benjamin Laenen, Marco Fracassetti, Aurélie Désamoré, Clément Lafon-Placette, Kim A. Steige, et al. 2019. Genetic basis and timing of a major mating system shift in *Capsella*. *New Phytologist* 224 (1): 505–517. https://doi.org/10.1111/nph.16035.

Backman, Talia, Sergio M Latorre, Efthymia Symeonidi, Artur Muszyński, Ella Bleak, Lauren Eads, Paulina I Martinez-Koury, et al. 2024. A phage tail-like bacteriocin suppresses competitors in metapopulations of pathogenic bacteria. *Science* 384 (6701): eado0713. https://doi.org/10.1126/science.ado071a.

Bakker, Erica G., Christopher Toomajian, Martin Kreitman, and Joy Bergelson. 2006. A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *The Plant Cell* 18 (8): 1803–1818. https://doi.org/10.1105/tpc.106.042614.

Bao, Zhigui, and Detlef Weigel. 2025. Complexity welcome: pangenome graphs for comprehensive population genomics. *Quantitative Plant Biology,* 1–34. https://doi.org/10.1017/qpb.2025.10028.

Barragan, A Cristina, and Detlef Weigel. 2021. Plant NLR diversity: the known unknowns of pan-NLRomes. *The Plant Cell* 33 (4): 814–831. https://doi.org/10.1093/plcell/koaa002.

Barton, N. H., and B. Charlesworth. 1998. Why sex and recombination? *Science* 281 (5385): 1986–1990. https://doi.org/10.1126/science.281.5385.1986.

Begun, DJ, and Charles F Aquadro. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* 365 (6446): 548–550. https://doi.org/10.1038/365548a0.

Bergelson, Joy, Martin Kreitman, Eli A. Stahl, and Dacheng Tian. 2001. Evolutionary dynamics of plant *R*-genes. *Science* 292 (5525): 2281–2285. https://doi.org/10.1126/science.1061337.

Bomblies, Kirsten, and Detlef Weigel. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* 8 (5): 382–393. https://doi.org/10.1038/nrg2082.

Boots, M, A White, A Best, and R. Bowers. 2014. How specificity and epidemiology drive the coevolution of static trait diversity in hosts and parasites. *Evolution* 68 (6): 1594–1606. https://doi.org/10.1111/evo.12393.

Bourgeois, Yann X. C., and Ben H. Warren. 2021. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular Ecology* 30 (23): 6036–6071. https://doi.org/10.1111/mec.15989.

Brown, James K. M., and Aurélien Tellier. 2011. Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology* 49 (Volume 49, 2011): 345–367. https://doi.org/10.1146/annurev-phyto-072910-095301.

Burdon, Jeremy J, and Anna-Liisa Laine. 2019. *Evolutionary dynamics of plant-pathogen interactions.* Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781108625517.

Campos, José L., Daniel L. Halligan, Penelope R. Haddrill, and Brian Charlesworth. 2014. The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. *Molecular Biology and Evolution* 31 (4): 1010–1028. https://doi.org/10.1093/molbev/msu056.

Charlesworth, Brian, and Deborah Charlesworth. 2010. *Elements of evolutionary genetics.* New York: Greenwood Village, Colorado: Roberts & Co. ISBN: 9780981519425.

Charlesworth, Brian, and Jeffrey D. Jensen. 2021. Effects of selection at linked sites on patterns of genetic variability. *Annual Review of Ecology, Evolution, and Systematics* 52 (1): 177–197. https://doi.org/10.1146/annurev-ecolsys-010621-044528.

Cheng, Haoyu, Gregory T Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18 (2): 170–175. https://doi.org/10.1038/s41592-020-01056-5.

Choi, Kyuha, Carsten Reinhard, Heïdi Serra, Piotr A. Ziolkowski, Charles J. Underwood, Xiaohui Zhao, Thomas J. Hardcastle, et al. 2016. Recombination rate heterogeneity within *Arabidopsis* disease resistance genes. *PLOS Genetics* 12 (7): e1006179. https://doi.org/10.1371/journal.pgen.1006179.

Clark, James S., Mike Dietze, Sukhendu Chakraborty, Pankaj K. Agarwal, Ines Ibanez, Shannon LaDeau, and Mike Wolosin. 2007. Resolving the biodiversity paradox. *Ecology Letters* 10 (8): 647–659. https://doi.org/10.1111/j.1461-0248.2007.01041.x.

Dawkins, Richard, and John Richard Krebs. 1979. Arms races between and within species. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 205 (1161): 489–511. https://doi.org/10.1098/rspb.1979.0081.

De Meaux, J., I. Cattan-Toupance, C. Lavigne, T. Langin, and C. Neema. 2003. Polymorphism of a complex resistance gene candidate family in wild populations of common bean (*Phaseolus vulgaris*) in Argentina: comparison with phenotypic resistance polymorphism. *Molecular Ecology* 12 (1): 263–273. https://doi.org/10.1046/j.1365-294X.2003.01718.x.

Dodds, Peter N., and John P. Rathjen. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* 11 (8): 539–548. https://doi.org/10.1038/nrg2812.

Dutheil, Julien Y., ed. 2020. *Statistical population genomics.* Methods in Molecular Biology. Springer Nature. https://doi.org/10.1007/978-1-0716-0199-0.

Fierst, Janna L., and Patrick C. Phillips. 2015. Modeling the evolution of complex genetic systems: the gene network family tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 324 (1): 1–12. https://doi.org/10.1002/jez.b.22597.

Fox, Emma A, Alison E Wright, Matteo Fumagalli, and Filipe G Vieira. 2019. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* 35 (19): 3855–3856. https://doi.org/10.1093/bioinformatics/btz200.

Freh, Matthias, Jinlan Gao, Morten Petersen, and Ralph Panstruga. 2021. Plant autoimmunity—fresh insights into an old phenomenon. *Plant Physiology* 188 (3): 1419–1434. https://doi.org/10.1093/plphys/kiab590.

Garrison, Erik, Andrea Guarracino, Simon Heumos, Flavia Villani, Zhigui Bao, Lorenzo Tattini, Jörg Hagmann, et al. 2024. Building pangenome graphs. *Nat. Methods* 21 (11): 2008–2012. https://doi.org/10.1038/s41592-024-02430-3.

Gos, Gesseca, Tanja Slotte, and Stephen I. Wright. 2012. Signatures of balancing selection are maintained at disease resistance loci following mating system evolution and a population bottleneck in the genus *Capsella*. *BMC Evolutionary Biology* 12 (1): 152. https://doi.org/10.1186/1471-2148-12-152.

Grandaubert, Jonathan, Julien Y. Dutheil, and Eva H. Stukenbrock. 2019. The genomic determinants of adaptive evolution in a fungal pathogen. *Evolution Letters* 3 (3): 299–312. https://doi.org/10.1002/evl3.117.

Hanlon, Vincent C. T., Alex Cagan, and Sebastian Eves-van den Akker. 2025. How and when organisms edit their own genomes. *Nat Genet* 57:1823–1834. https://doi.org/10.1038/s41588-025-02230-1.

Hartl, Daniel L., and Andrew G. Clark. 2023. *Principles of population genetics.* 4th ed. Oxford University Press Inc. ISBN: 9780197778739.

Hill, W. G., and Alan Robertson. 2007. The effect of linkage on limits to artificial selection. *Genetics Research* 89 (5): 311–336. https://doi.org/10.1017/S001667230800949X.

Hörger, Anja C., Muhammad Ilyas, Wolfgang Stephan, Aurélien Tellier, Renier A. L. van der Hoorn, and Laura E. Rose. 2012. Balancing selection at the tomato RCR3 guardee gene family maintains variation in strength of pathogen defense. *PLOS Genetics* 8 (7): e1002813. https://doi.org/10.1371/journal.pgen.1002813.

Igolkina, Anna A., Sebastian Vorbrugg, Fernando A. Rabanal, Hai-Jun Liu, Haim Ashkenazy, Aleksandra E. Kornienko, Joffrey Fitz, et al. 2025. A comparison of 27 *Arabidopsis thaliana* genomes and the path toward an unbiased characterization of genetic polymorphism. *Nature Genetics* 57 (9): 2289–2301. ISSN: 1546-1718. https://doi.org/10.1038/s41588-025-02293-0.

Innan, Hideki. 2003. The coalescent and infinite-site model of a small multigene family. *Genetics* 163 (2): 803–810. ISSN: 1943-2631. https://doi.org/10.1093/genetics/163.2.803.

Innan, Hideki, and Fyodor Kondrashov. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* 11 (2): 97–108. https://doi.org/10.1038/nrg2689.

Jiao, Wen-Biao, Vipul Patel, Jonas Klasen, Fang Liu, Petra Pecinkova, Marina Ferrand, Isabelle Gy, et al. 2021. The evolutionary dynamics of genetic incompatibilities introduced by duplicated genes in *Arabidopsis thaliana*. *Molecular Biology and Evolution* 38 (4): 1225–1240. https://doi.org/10.1093/molbev/msaa306.

Johri, Parul, Charles F. Aquadro, Mark Beaumont, Brian Charlesworth, Laurent Excoffier, Adam Eyre-Walker, Peter D. Keightley, et al. 2021. *Recommendations for improving statistical inference in population genomics.* https://doi.org/10.1101/2021.10.27.466171.

Karasov, Talia L., Joel M. Kniskern, Liping Gao, Brody J. DeYoung, Jing Ding, Ullrich Dubiella, Ruben O. Lastra, et al. 2014. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* 512 (7515): 436–440. https://doi.org/10.1038/nature13439.

Keightley, Peter D, and Adam Eyre-Walker. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177 (4): 2251–2261. ISSN: 1943-2631. https://doi.org/10.1534/genetics.107.080663.

Koenig, Daniel, Jörg Hagmann, Rachel Li, Felix Bemm, Tanja Slotte, Barbara Neuffer, Stephen I Wright, and Detlef Weigel. 2019. Long-term balancing selection drives evolution of immunity genes in capsella. *eLife* 8:e43606. https://doi.org/10.7554/eLife.43606.

Korfmann, Kevin, Thibaut Paul Patrick Sellinger, Fabian Freund, Matteo Fumagalli, and Aurélien Tellier. 2024. Simultaneous Inference of Past Demography and Selection from the Ancestral Recombination Graph under the Beta Coalescent. *Peer Community Journal* 4: e33. https://doi.org/10.24072/pcjournal.397.

Kryazhimskiy, Sergey, and Joshua B Plotkin. 2008. The population genetics of dN/dS. *PLoS Genet.* 4 (12): e1000304.

Laflamme, Bradley, Marcus M. Dillon, Alexandre Martel, Renan N. D. Almeida, Darrell Desveaux, and David S. Guttman. 2020. The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science* 367 (6479): 763–768. https://doi.org/10.1126/science.aax4079.

Lee, Rachelle R.Q., and Eunyoung Chae. 2020. Variation patterns of NLR clusters in *Arabidopsis thaliana* genomes. *Plant Communications* 1 (4): 100089. https://doi.org/10.1016/j.xplc.2020.100089.

Li, Sai-Xi, Yang Liu, Yan-Mei Zhang, Jian-Qun Chen, and Zhu-Qing Shao. 2025. Convergent reduction of immune receptor repertoires during plant adaptation to diverse special lifestyles and habitats. *Nature Plants* 11 (2): 248–262. https://doi.org/10.1038/s41477-024-01901-x.

Li, Yuying, Qiong Wang, Huimin Jia, Kazuya Ishikawa, Ken-ichi Kosami, Takahiro Ueba, Atsumi Tsujimoto, et al. 2024. An NLR paralog Pit2 generated from tandem duplication of Pit1 fine-tunes Pit1 localization and function. *Nature Communications* 15:4610. https://doi.org/10.1038/s41467-024-48943-5.

Lian, Qichao, Victor Solier, Birgit Walkemeier, Stéphanie Durand, Bruno Huettel, Korbinian Schneeberger, and Raphael Mercier. 2022. The megabase-scale crossover landscape is largely independent of sequence divergence. *Nature Communications* 13 (1): 3828. ISSN: 2041-1723. https://doi.org/10.1038/s41467-022-31509-8. 10.1038/s41467-022-31509-8.

Lynch, Michael. 2007. *The origins of genome architecture.* Oxford University Press Inc. ISBN: 978-0-87893-484-3.

Michelmore, Richard W., and Blake C. Meyers. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* 8 (11): 1113–1130. https://doi.org/10.1101/gr.8.11.1113.

Milgroom, Michael G. 2023. *Biology of infectious disease: From molecules to ecosystems.* Springer Cham. ISBN: 978-3-031-38940-5. https://doi.org/10.1007/978-3-031-38941-2.

Otto, Moritz, Yichen Zheng, and Thomas Wiehe. 2022. Recombination, selection, and the evolution of tandem gene arrays. *Genetics* 221 (3): iyac052. https://doi.org/10.1093/genetics/iyac052.

Pouzet, Sylvain, and Arnaud Le Rouzic. 2024. Gene network topology drives the mutational landscape of gene expression. *bioRxiv,* https://doi.org/10.1101/2024.11.28.625874.

Pradeu, Thomas, Bart P H J Thomma, Stephen E Girardin, and Bruno Lemaitre. 2024. The conceptual foundations of innate immunity: taking stock 30 years later. *Immunity* 57 (4): 613–631. https://doi.org/10.1016/j.immuni.2024.03.007.

Prigozhin, Daniil M, and Ksenia V Krasileva. 2021. Analysis of intraspecies diversity reveals a subset of highly variable plant immune receptors and predicts their binding sites. *The Plant Cell* 33 (4): 998–1015. https://doi.org/10.1093/plcell/koab013.

Rautiainen, Mikko, Sergey Nurk, Brian P Walenz, Glennis A Logsdon, David Porubsky, Arang Rhie, Evan E Eichler, Adam M Phillippy, and Sergey Koren. 2023. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat. Biotechnol.* 41 (10): 1474–1482. https://doi.org/10.1038/s41587-023-01662-6.

Rhie, Arang, Shane A McCarthy, Olivier Fedrigo, Joana Damas, Giulio Formenti, Sergey Koren, Marcela Uliano-Silva, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 592 (7856): 737–746. https://doi.org/10.1038/s41586-021-03451-0.

Rose, Laura E, Richard W Michelmore, and Charles H Langley. 2007. Natural variation in the Pto disease resistance gene within species of wild tomato (*Lycopersicon*). II. Population genetics of Pto. *Genetics* 175 (3): 1307–1319. ISSN: 1943-2631. https://doi.org/10.1534/genetics.106.063602.

Sakai, Toshiyuki, Mauricio P Contreras, Claudia Martinez-Anaya, Daniel Lüdke, Sophien Kamoun, Chih-Hang Wu, and Hiroaki Adachi. 2024. The NRC0 gene cluster of sensor and helper NLR immune receptors is functionally conserved across asterid plants. *The Plant Cell* 36 (9): 3344–3361. ISSN: 1040-4651. https://doi.org/10.1093/plcell/koae154.

Sellinger, Thibaut Paul Patrick, Diala Abu Awad, Markus Moest, and Aurélien Tellier. 2021. Correction: inference of past demography, dormancy and self-fertilization rates from whole genome sequence data. *PLoS Genetics* 17 (4): e1009504. https://doi.org/10.1371/journal.pgen.1009504.

Sendrowski, Janek, and Thomas Bataillon. 2024. fastDFE: fast and flexible inference of the distribution of fitness effects. *Molecular Biology and Evolution* 41 (5): msae070. https://doi.org/10.1093/molbev/msae070.

Seong, Kyungyong, Eunyoung Seo, Kamil Witek, Meng Li, and Brian Staskawicz. 2020. Evolution of NLR resistance genes with noncanonical N-terminal domains in wild tomato species. *New Phytologist* 227 (5): 1530–1543. https://doi.org/10.1111/nph.16628.

Shalev, Or, Talia L Karasov, Derek S Lundberg, Haim Ashkenazy, Pratchaya Pramoj Na Ayutthaya, and Detlef Weigel. 2022. Commensal pseudomonas strains facilitate protective response against pathogens in the host plant. *Nat. Ecol. Evol.* 6 (4): 383–396. https://doi.org/10.1038/s41559-022-01673-7.

Silva-Arias, Gustavo A., Edeline Gagnon, Surya Hembrom, Alexander Fastner, Muhammad Ramzan Khan, Remco Stam, and Aurélien Tellier. 2025. Patterns of presence–absence variation of NLRs across populations of *Solanum chilense* are clade-dependent and mainly shaped by past demographic history. *New Phytologist* 245 (4): 1718–1732. https://doi.org/10.1111/nph.20293.

Stahl, Eli A., Greg Dwyer, Rodney Mauricio, Martin Kreitman, and Joy Bergelson. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature* 400 (6745): 667–671. https://doi.org/10.1038/23260.

Stam, Remco, Tetyana Nosenko, Anja C Hörger, Wolfgang Stephan, Michael Seidel, José M M Kuhn, Georg Haberer, and Aurelien Tellier. 2019. The de novo reference genome and transcriptome assemblies of the wild tomato species *Solanum chilense* highlights birth and death of NLR genes between tomato species. *G3 Genes|Genomes|Genetics* 9 (12): 3933–3941. https://doi.org/10.1534/g3.119.400529.

Strütt, Stefan, Thibaut Sellinger, Sylvain Glémin, Aurélien Tellier, and Stefan Laurent. 2023. Joint inference of evolutionary transitions to self-fertilization and demographic history using whole-genome sequences. *eLife* 12:e82384. https://doi.org/10.7554/eLife.82384.

Sutherland, Chandler A, Daniil M Prigozhin, J Grey Monroe, and Ksenia V Krasileva. 2024. High allelic diversity in *Arabidopsis* NLRs is associated with distinct genomic features. *EMBO Reports* 25 (5): 2306–2322. https://doi.org/10.1038/s44319-024-00122-9.

Teasdale, Luisa C., Kevin D. Murray, Max Collenberg, Adrian Contreras-Garrido, Theresa Schlegel, Leon van Ess, Justina Jüttner, et al. 2025. Pangenomic context reveals the extent of intraspecific plant NLR evolution. *Cell Host & Microbe* 33:1291. https://doi.org/10.1016/j.chom.2025.07.011.

Tellier, Aurélien, and James K M Brown. 2007. Stability of genetic polymorphism in host–parasite interactions. *Proceedings of the Royal Society B: Biological Sciences* 274 (1611): 809–817. https://doi.org/10.1098/rspb.2006.0281.

Tellier, Aurélien, Stefany Moreno-Gámez, and Wolfgang Stephan. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics: genomic signatures of coevolution. *Evolution,* 2211–2224. https://doi.org/10.1111/evo.12427.

Thompson, John N. 2005. *The geographic mosaic of coevolution.* University of Chicago Press. ISBN: 978-0-226-11869-7. https://doi.org/10.7208/chicago/9780226118697.001.0001.

Thornton, Kevin R. 2007. The neutral coalescent process for recent gene duplications and copy-number variants. *Genetics* 177 (2): 987–1000. ISSN: 1943-2631. https://doi.org/10.1534/genetics.107.074948.

Tian, Dacheng, Hitoshi Araki, Eli Stahl, Joy Bergelson, and Martin Kreitman. 2002. Signature of balancing selection in *Arabidopsis*. *Proceedings of the National Academy of Sciences* 99 (17): 11525–11530. https://doi.org/10.1073/pnas.172203599.

Van de Weyer, Anna-Lena, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D.G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. 2019. A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* 178 (5): 1260–1272.e14. https://doi.org/10.1016/j.cell.2019.07.038.

Van der Hoorn, Renier A.L, Pierre J.G.M De Wit, and Matthieu H.A.J Joosten. 2002. Balancing selection favors guarding resistance proteins. *Trends in Plant Science* 7 (2): 67. https://doi.org/10.1016/S1360-1385(01)02188-4.

Van Valen, Leigh. 1973. A new evolutionary law. *Evolutionary theory* 1:1–30.

Wagner, Andreas. 1996. Does evolutionary plasticity evolve? *Evolution* 50 (3): 1008–1023. https://doi.org/10.1111/j.1558-5646.1996.tb02342.x.

Wei, Kai, Saida Sharifova, Xiaoyun Zhao, Neelima Sinha, Hokuto Nakayama, Aurélien Tellier, and Gustavo A. Silva-Arias. 2024. Evolution of gene networks underlying adaptation to drought stress in the wild tomato *Solanum chilense*. *Molecular Ecology* 33 (21): e17536. https://doi.org/10.1111/mec.17536.

Wei, Kai, Gustavo A. Silva-Arias, and Aurélien Tellier. 2023. Selective sweeps linked to the colonization of novel habitats and climatic changes in a wild tomato species. *New Phytologist* 237 (5): 1908–1921. https://doi.org/10.1111/nph.18634.

Weiner, Benjamin G., Hanna Märkle, Eric Laderman, Choghag Demirjian, and Joy Bergelson. 2025. A physical model links structure and function in the plant immune system. *Proceedings of the National Academy of Sciences* 122 (24): e2502872122. https://doi.org/10.1073/pnas.2502872122.

Wiuf, Carsten, and Jotun Hein. 1997. On the number of ancestors to a DNA sequence. *Genetics* 147 (3): 1459–1468. https://doi.org/10.1093/genetics/147.3.1459.

Woolhouse, Mark E. J., Joanne P. Webster, Esteban Domingo, Brian Charlesworth, and Bruce R. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* 32 (4): 569–577. https://doi.org/10.1038/ng1202-569.

Wu, Chih-Hang, Ahmed Abd-El-Haliem, Tolga O. Bozkurt, Khaoula Belhaj, Ryohei Terauchi, Jack H. Vossen, and Sophien Kamoun. 2017. NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences* 114 (30): 8113–8118. https://doi.org/10.1073/pnas.1702041114.

Wu, Qiong, Ting-Shen Han, Xi Chen, Jia-Fu Chen, Yu-Pan Zou, Zi-Wen Li, Yong-Chao Xu, and Ya-Long Guo. 2017. Long-term balancing selection contributes to adaptation in *Arabidopsis* and its relatives. *Genome Biology* 18. https://doi.org/10.1186/s13059-017-1342-8.

Yang, Yu, Nak Hyun Kim, Volkan Cevik, Pierre Jacob, Li Wan, Oliver J. Furzer, and Jeffery L. Dangl. 2022. Allelic variation in the Arabidopsis TNL CHS3/CSA1 immune receptor pair reveals two functional cell-death regulatory modes. *Cell Host & Microbe* 30:1701–1716. https://doi.org/10.1016/j.chom.2022.09.013.

Živković, Daniel, Sona John, Mélissa Verin, Wolfgang Stephan, and Aurélien Tellier. 2019. Neutral genomic signatures of host-parasite coevolution. 19:230. https://doi.org/10.1186/s12862-019-1556-3.

Zou, Cheng, Melissa D. Lehti-Shiu, Françoise Thibaud-Nissen, Tanmay Prakash, C. Robin Buell, and Shin-Han Shiu. 2009. Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. *Plant Physiology* 151 (1): 3–15. https://doi.org/10.1104/pp.109.140632.

Box 1: Ten open ecological and evolutionary questions in NLR gene research

We formulate the following questions as guidelines arising from the studies mentioned in the text (which are mainly based on *A. thaliana*).

1. What is the mutation rate at NLR genes in plant species?
2. What are the recombination rates at NLR genes in selfing versus outcrossing species?
3. Which polymorphism patterns are found at NLR genes in tandem arrangements in species with higher or lower linkage disequilibrium (selfing or outcrossing)?
4. Is the extent of PAV and CNV of NLR genes different in plant species with different life-history traits and ecological contexts (the latter defining the rates of genetic drift, mutation, recombination and gene duplication/deletion)?
5. What is the extent and occurrence of pseudogenization in NLR genes? How does this relate to plant life-history traits and the ecological context?
6. What is the number, structure, connectivity, size, and redundancy of NLR gene networks in different plant species?
7. Do NLR gene network topologies depend on the ecological context?
8. Is the position of individual NLR genes in the networks related to their recognition of pathogens, mutualists and/or regulation of interactions with the (soil, leaf,…) microbiome?
9. What type of selection occurs at which gene in the NLR gene networks?
10. Does the process of NLR evolution follow a punctuated equilibrium (saltationary) model of evolution with the speed of evolution (and diversity) being discretely distributed across loci? Alternatively, is the distribution of evolutionary rates more continuous across NLRs?

Figures

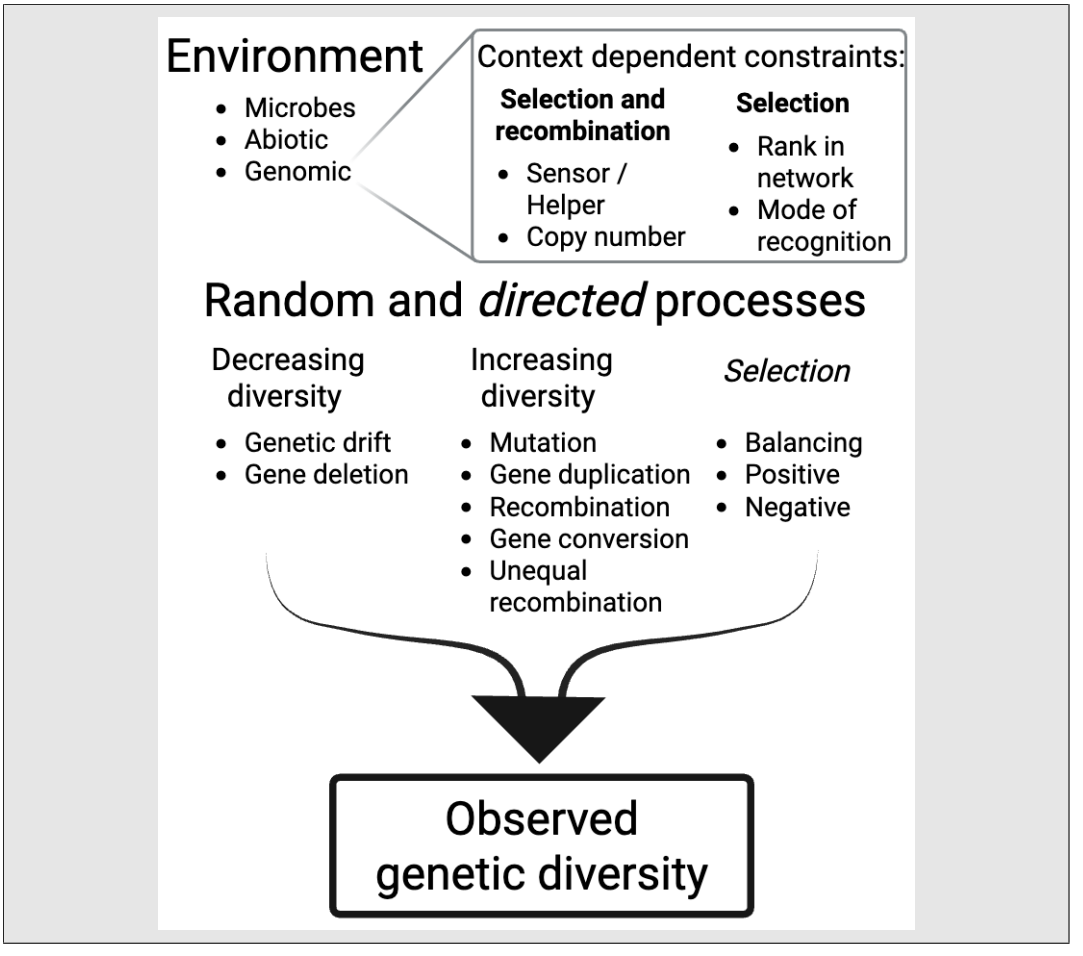

Figure 1. Overview of the different evolutionary forces and constraints driving NLR gene family evolution. The environment of a given NLR gene comprises the external abiotic and biotic factors, as well as the genomic space of the considered host plant, including the organization (number and localization) of NLR (homologs and paralogs).

Table: Timescales of epidemiology, population genomics and evolution

Table 1. Timescales of epidemiology, population genomics and evolution

Timescale [years]	Process	Data used	Theoretical framework
one to a few	Epidemiological dynamics	Disease prevalence and severity	Epidemiology
up to a few thousand	Ecological-evolutionary dynamics (agriculture: boom-and-bust cycles)	Within-species polymorphism	Population genetics (micro-evolution)
millions	Evolution of genome structure and genome space	Between-species divergence and synteny	Macro-evolution

Table: Summary of population genomics analyses to interrogate the evolution of NLR genes and clusters

To keep the list of references manageable, we refrain from citing all publications and method papers referenced in the table. We direct readers to textbooks (Wiuf and Hein 1997, Charlesworth and Charlesworth 2010, Hartl and Clark 2023) and the reference manual of state-of-the-art population genomics methods (Dutheil 2020). Recommendations to perform population genomics analyses rigorously are found in Johri et al. 2021, Bourgeois and Warren 2021 and in ad hoc chapters of Dutheil 2020. Software recommendations can be found in Bourgeois and Warren 2021 and we only indicate here recent additions: ngsLD (Fox et al. 2019), eSMC2 (Sellinger et al. 2021), and SMβC (Korfmann et al. 2024).

Table 2. Summary of population genomics analyses to interrogate the evolution of NLR genes and clusters

Type of analysis	Data required	Computation of measure (statistics)	General description and interpretations
Genetic diversity for SNPs or structural variants	Genomic data for several individuals (SNP or SV calls) with SV coded as PAV (bi-allelic state)	Pairwise nucleotide diversity (π), Watterson Theta (Θw)	Diversity per gene is a function of the effective population size (N_e), determining the effect of genetic drift and mutation input, past demographic events, and selective effects. Nucleotide diversity is positively correlated with local recombination, as high recombination decreases the negative effect of linked (positive or negative) selection on nearby variants.

Table 2. Summary of population genomics analyses to interrogate the evolution of NLR genes and clusters

Type of analysis	Data required	Computation of measure (statistics)	General description and interpretations
Recombination		Linkage disequilibrium (LD) (r^2 is a correlation coefficient between pairs of loci)	Linkage disequilibrium is a measure of meiotic recombination along the genome. Effective recombination depends on the effective population size (N_e) and per-site recombination rate which may vary along the genome (for example lower in centromeric regions). LD measure should indicate also if neighboring genes (e.g. tandem NLRs) are linked in their evolution by chance.
Neutrality tests	Genomic data for several individuals (SNP calls) with outgroup species for SNP polarization.	Tajima's D, Fu and Li's D, Fay and Wu's H	Tajima's D and other neutrality tests measure the deviation of the Site Frequency Spectrum (SFS), that is the distribution of polarized allele frequencies per genomic region (or locus), to the expected neutral distribution. Selective processes acting at peculiar loci are seen as outliers for these statistics compared to the genome-wide distribution. Past demographic events and variation in recombination rates along the genome should be counted when scanning for selection outliers or nearby loci (e.g. tandem NLRs).
Gene conversion	Genomic data for several individuals at paralogs (SNP calls) with outgroup species for SNP polarization.	Joint SFS for both paralogs	The joint SFS computation for shared and private allele frequencies between paralogs allow to infer the occurrence of gene conversion. This can allow to derive threshold and expected distribution for neutrality tests to assess the occurrence of positive or balancing selection at paralogs.

Table 2. Summary of population genomics analyses to interrogate the evolution of NLR genes and clusters

Type of analysis	Data required	Computation of measure (statistics)	General description and interpretations
Neutrality tests with outgroup	Genomic data for several individuals (SNP calls) with outgroup species for SNP polarization at coding regions.	πN, πS, ratio πN /πS, McDonald-Kreitman (MK) test, Hudson-Kreitman-Aguade (HKA) test, ω, Distribution of Fitness Effects (DFE) estimation	Nucleotide diversity at synonymous (πS) and non-synonymous (πN) sites and the ratio (πN /πS) measure selective or relaxed constraints at coding regions. These are influenced by the effective population size (N_e) and drift, past demographic events as well as recombination rate and selective processes (negative and positive selection). The MK and HKA tests do compare polymorphism versus substitutions at coding regions, assuming synonymous sites are nearly neutral. An extension of the MK test is the inference of the percentage of sites under positive selection (ω) and the distribution of fitness effects which is inferred based on the SFS of synonymous and non-synonymous sites. The strength of genetic drift, the amount of recombination and effect of linked selection are key determinants of the ω and DFE estimates.
Selection across phylogeny	Genomic data for several species (coding regions)	KA (dN), KS (dS), ratio KA/KS (dN/dS)	The substitution rates at synonymous (KS or dS), non-synonymous (KA, dN) and their ratio (KA/KS or dN/dS) indicate long-term deviations from the neutral evolution at coding genes. These are computed over phylogenetic timescales when comparing several species.
Genetic differentiation between populations (population structure)	Genomic data for several individuals (SNP or SV calls) with SV coded as PAV (bi-allelic state)	Indices of fixation or population differentiation: for SNPs FST, Jost's D, dXY, and for SV (FST or VST).	Population structure and gene flow between populations, as well as genetic drift and selective events (local adaptation), influence the values of measures of differentiation. Outlier genes for high differentiation compared to the genome distribution may indicate positive selection for local adaptation, while low differentiation compared to the genome-wide distribution may indicate balancing selection. The footprints of population differentiation depend on the rate of recombination.

Table 2. Summary of population genomics analyses to interrogate the evolution of NLR genes and clusters

Type of analysis	Data required	Computation of measure (statistics)	General description and interpretations
Inference of neutral demographic and spatial processes	Genomic data for several individuals (SNP or SV calls) with SV coded as PAV (bi-allelic state)	Based on nucleotide diversity, LD, and population differentiation.	A wide range of methods build on nucleotide diversity and LD distribution along the genome to infer the past demographic and population structure of a population. Based on such inference, one can predict the threshold for outlier loci for various neutrality tests to detect selective events. Such thresholds should account for variation in recombination rate and possible linked selection effects.
Genome scans for positive selection	Genomic data for several individuals (SNP calls) with outgroup species for SNP polarization	Based on nucleotide diversity and linkage disequilibrium per genomic region	These methods scan the genome to detect genes/regions under positive selection. The threshold for outlier regions is dset based on the inference of past demographic events and population structure.