

Part B2: *The Project Proposal*

Section a. State of the art and objectives

Motivation and goals of IMMUNEMESIS

A fundamental question in biology is how multicellular organisms **distinguish self and non-self**. If they were not able to make this distinction, animals and plants could not detect and respond to pathogens without triggering immune reactions directed against their own cells. This requirement places constraints on the diversification of pathogen recognition systems, resulting in conflicts between effective detection of enemies, adaptive changes in the cellular machinery and mating with divergent genotypes from the same species. In plants, there is generally a **trade-off between immunity and growth**, and priming of the immune system to respond to pathogens is often associated with impaired development¹. Although this is a widely observed phenomenon, we know little about the genetic basis for the connection between the two traits, nor how the trade-off between preparedness for pathogen attack and unintended collateral damage is managed.

In plants, there are many examples of spontaneous activation of the immune system in F₁ hybrids, due to **epistatic, non-additive interactions** between (mostly unknown) genes contributed by the different parents². An overt sign of such autoimmunity is often leaf necrosis, and **hybrid necrosis** has been described in both wild species and crops³. The first case to be studied molecularly was in tomato, where Jonathan Jones and colleagues found that introgression of a disease resistance locus from a wild species had been accompanied by simultaneous introduction of a second, unlinked locus that prevented spontaneous activity of the resistance gene⁴. Importantly, autoimmunity does not have to be detrimental. On the contrary, plants with signs of autoimmunity are often (though not always) more resistant to pathogens⁵⁻⁹, and mutants with **autoimmune phenotypes** are resources in breeding for disease resistance^{10,11}.

Another common phenomenon in F₁ progeny from two inbred parents is hybrid vigor, or **heterosis**. This has been exploited for many decades in corn, which is naturally outcrossing, but the breeding of commercial F₁ hybrids is gaining traction in many selfing crops, including rice, wheat, soybean, cotton, tomato and canola^{12,13}. Thus, it is also of practical importance to understand how the immune system is affected when divergent genotypes are combined. In the model plant *Arabidopsis thaliana*, we have discovered random crosses between natural strains (accessions) that result in F₁ hybrid necrosis. The syndrome spans a range of severity, which can be influenced by the environment, gene dosage and genetic background, which is one of the reasons why we believe that the cases described so far represent only the tip of the iceberg. We hypothesize that **subtle epistatic interactions** that affect both the state of the plant immune system and growth are ubiquitous. I therefore propose a comprehensive research program to dissect non-additive genetic interactions with effects on plant growth and health. Our specific aims are:

- Produce a systematic map of dominant hybrid necrosis loci in *A. thaliana*
- Discover the species-wide spectrum of epistatic interactions priming the *A. thaliana* immune system
- Compare systematic maps of dominant and recessive epistatic interactions
- Contrast epistatic interactions in inbreeding and outcrossing *Arabidopsis*
- Link diversity in microbial communities and the immune system in natural plant populations

In the following, I will first introduce our model system, *A. thaliana* and its relatives, and the plant immune system. Next, I will describe how our previous efforts have laid the foundation for IMMUNEMESIS.

Genomics-enabled genetics in the model plant *Arabidopsis thaliana* and its relatives

Over the last decade, *A. thaliana* has become a **prime model for the study of genetic variation**. The availability of naturally inbred strains enables repeated phenotyping of the same, often locally adapted genotypes under diverse conditions¹⁴. To accelerate the discovery of DNA sequence variants that affect phenotypic differences, we, in collaboration with others, have pioneered large-scale studies of genotypic variation. Our first major effort in this area was the resequencing of 20 diverse accessions using ultra-high density microarrays¹⁵. This led to the **first haplotype map outside mammals**, and the common single nucleotide polymorphisms (SNPs) from this project have now been typed in well over a thousand freely available accessions, making it possible for anyone to conduct genome-wide association studies (GWAS)^{16,17}. A logical extension of our SNP discovery efforts was to generate more **comprehensive whole-genome information** for a large number of accessions. My lab was an early adopter of Illumina sequencing technology, and we were the first to use this platform for investigating sequence variation in plants. We developed our own analysis tools, including approaches for the detection of copy number variation and for

targeted assembly of sequences absent from the reference, concurrently with similar efforts for human genomes¹⁸. Based on these early successes, we advocated a 1001 Genomes Project for *A. thaliana*^{19,20} (<http://1001genomes.org>). Last year, we published the results from the first major phase of this endeavor, reporting on 80 genomes from eight geographically diverse populations²¹. To better understand patterns of sequence variation in *A. thaliana*, I also took the lead in obtaining high-quality genome sequences for two *A. thaliana* relatives, the outcrosser *Arabidopsis lyrata*²², and *Capsella rubella*, which made the transition to selfing much more recently than *A. thaliana*^{23,24}.

The plant immune system

Plants are under relentless attack by diverse predators, from microbes (bacteria, fungi and oomycetes) to animals (in particular insects and nematodes). In addition to physical defenses, plants have evolved a sophisticated two-tiered immune system, similar to vertebrates. The first tier relies on the recognition of conserved signatures associated with essential molecules found in broad classes of pathogens. These are known as pathogen- or microbe-associated molecular patterns (PAMPs/MAMPs). In addition, pathogen attack can lead to the release of molecules from the plant itself, and these endogenous danger signals have been dubbed damage-associated molecular patterns (DAMPs). Both PAMPs and DAMPs are detected by pattern recognition receptors (PRRs). Upon PRR activation, a pathogen suppression program is triggered, with a MAP kinase cascade connecting the receptors to nuclear WRKY transcription factors (Fig. 1). This first arm of defense is known as **PAMP/DAMP-triggered immunity (PTI)**^{25,26}.

To overcome PTI, pathogens inject countermeasures in the form of **effectors** into the apoplastic space or directly into host cells; the effectors inhibit digestive enzymes secreted by plant cells or interfere with signal transduction. Unlike proteins that are recognized as PAMPs, effectors are not necessarily essential for pathogen viability. The action of effectors is in turn detected by the second arm of the plant immune system, which leads to **effector-triggered immunity (ETI)**²⁷ (Fig. 1). Central roles in ETI are played by the products of **resistance (R)** genes, most of which encode leucine-rich repeat proteins with a nucleotide-binding site (NB-LRR proteins). Proteins with a similar organization are important in animal innate immunity. Perhaps counter-intuitively, R proteins often do not interact directly with effector proteins. Rather, they monitor the conformation or integrity of endogenous plant proteins that are targeted by the effectors. The recognition of altered self by an R protein results in ETI responses, which are qualitatively similar to PTI responses, but more robust and often associated with cell death, known as **hypersensitive response (HR)**, which is visible as **leaf necrosis**^{26,27}. To bypass ETI, pathogens must jettison effector genes, which reduces their pathogenicity. The waxing and waning levels of immunity under this scenario have been referred to as the zigzag mode of defense²⁷. The distinction between PAMP/DAMP and effector recognition has recently become blurred²⁸, but in general R genes appear to be more polymorphic within species than PRR genes. This was initially deduced from genetic analyses, and has since been confirmed by genome-wide analyses, including our resequencing studies in *A. thaliana*^{15,21,29}.

Autoimmunity in hybrids

When independently diverging genomes meet in hybrids, the differences that have accumulated over time can have unintended, detrimental consequences. In the 1930s and 1940s, Dobzhansky and Muller proposed

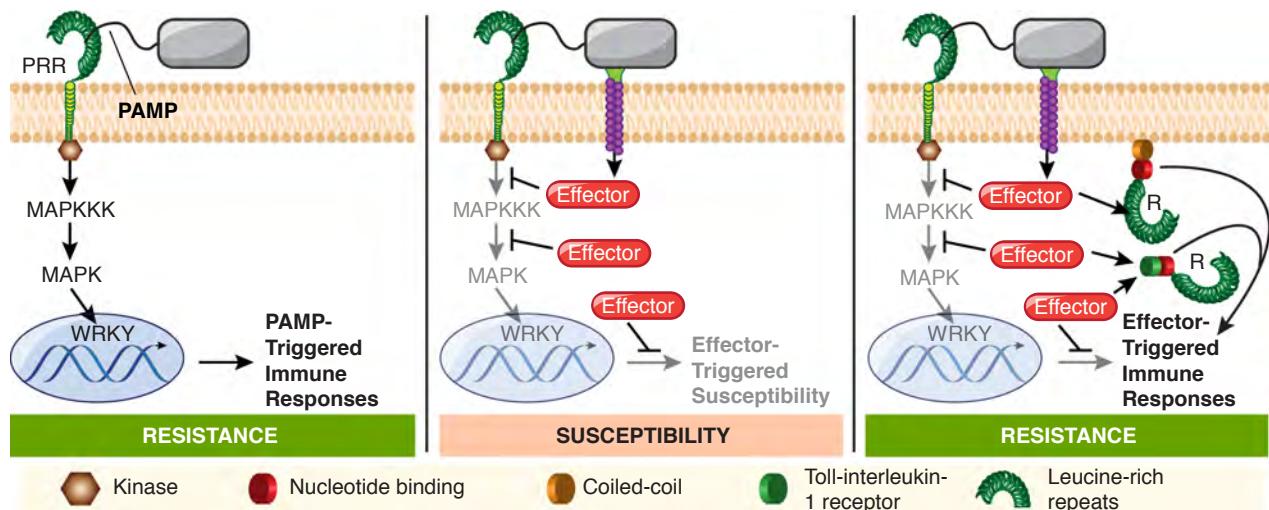


Fig. 1. Diagram of the different phases of PTI and ETI (after ref. 30). PRR – pattern recognition receptor; R – R protein.

models to describe how such incompatibilities can accumulate without ill effects in the parental lineages^{31,32}. Molecular genetic analyses have recently begun to reveal the causal genes underlying such interactions. One common finding has been that most **evolve faster** than the genome average, suggesting adaptation as an important driving force³³.

There is a rich literature on postzygotic incompatibilities in plants, which are at least as common as prezygotic barriers³⁴. Particularly prevalent causes of hybrid sterility are differences in ploidy. In addition, incompatibilities between nuclear and organellar genomes are often expressed as cytoplasmic male sterility (CMS)³⁵. Another type of sometimes deleterious **epistatic interaction** in plants is **hybrid necrosis** or weakness; until recently, it has received little attention in mainstream evolutionary literature³. It stands out among the various hybrid syndromes, because it is **common**, it is **phenotypically similar in many different species**, it has been found both within and between species, and it has generally **simple genetics**. While most often reported in the F₁, some cases of hybrid necrosis manifest themselves only in subsequent generations. The hybrids express **disease symptoms**, such as leaf necrosis, independently of pathogen attack. An intuitive interpretation is that a gene product contributed by one parent is inappropriately recognized as foreign by a gene product contributed by the other parent.

We have recently discovered that *A. thaliana* is an excellent choice for understanding epistatic interactions that activate the immune system. We found that about 2% of random crosses among accessions resulted in HR-like symptoms in the F₁ indicative of hybrid necrosis² (Fig. 2). The first *A. thaliana* hybrid necrosis gene we cloned, *DANGEROUS MIX1 (DMI)*, turned out to encode an **NB-LRR protein** with an N-terminal TIR (Toll-Interleukin-1 Receptor) domain². The Uk-3 accession carries a *DMI* allele that is incompatible with the Uk-1 allele at the unlinked *DM2* locus. Like *DMI*, *DM2* maps to a cluster of *NB-LRR* genes (Fig. 3). These observations mirror results from **tomato, lettuce and rice**, where hybrid necrosis loci encode either NB-LRR type R proteins, or R protein interactors^{4,36,37}.

To assess how widely variation in R proteins contributes to hybrid necrosis in *A. thaliana*, we have mapped additional genes responsible for four cases of different severity². We discovered a case that also involves a *DM2* protein, but with a different interactor, the *DM3* prolyl aminopeptidase (Figs. 3, 4). The interaction of *DM2* with *DM1* and *DM3* proteins can be reconstituted in *Nicotiana benthamiana*, confirming the results from knockdowns and retransformation in *A. thaliana* (Fig. 3). While we do not yet know what the endogenous function of *DM3* is, other proteases have roles in defense^{38,39}. The *DM2* locus appears to be a hotspot for generating autoimmune interactions, since an F₂ hybrid necrosis case maps to *DM2* as well, with the interacting *SRF3* gene encoding a receptor-like kinase that has been implicated in pathogen defense^{40,41} (Fig. 4). In addition, a dominant, EMS-induced mutation in a lyase gene can cause autoimmunity in combination with a specific allele at the *DM2* locus⁴².

In a third cross, we found that a particular allele of *RPW8*, which encodes an **atypical R protein** with a coiled-coil domain⁴³, likely interacts with a nucleoporin gene (Fig. 4). Finally, in a fourth cross, different alleles at the same locus, *ACD6*, interact in hybrids to activate the immune system. This case is particularly interesting because the F₁ hybrids are mildly affected (“class I”, Fig. 2) and can easily set seeds. Like other hybrid necrosis individuals, they are more resistant to pathogens, which

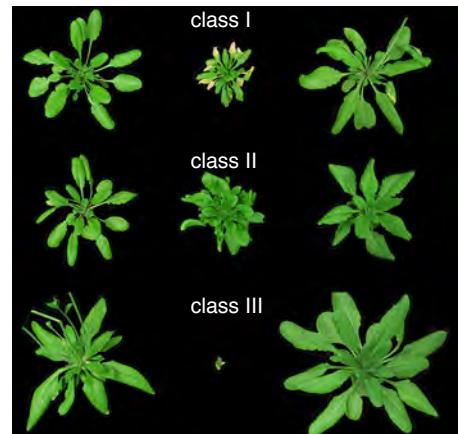


Fig. 2. Hybrid necrosis in *A. thaliana*. F₁ hybrids are shown between parents.

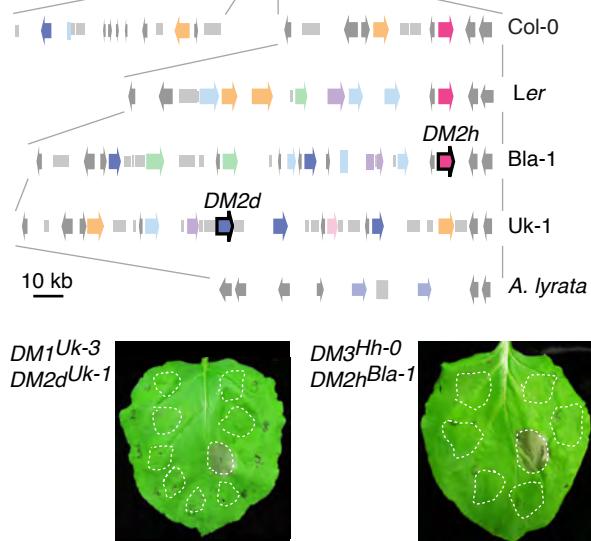


Fig. 3. Top, The *DM2* cluster of *NB-LRR* genes in *A. thaliana* accessions and in *A. lyrata*. *NB-LRR* genes are color-coded to indicate sequence similarity within and between accessions. Causal genes in *Bla-1* and *Uk-1* are highlighted. Bottom, Reconstitution of *DM1/DM2* and *DM3/DM2* interactions in *N. benthamiana*. Various controls, such as closely related *DM2* proteins or P-loop mutants, did not produce HR-like necrosis.

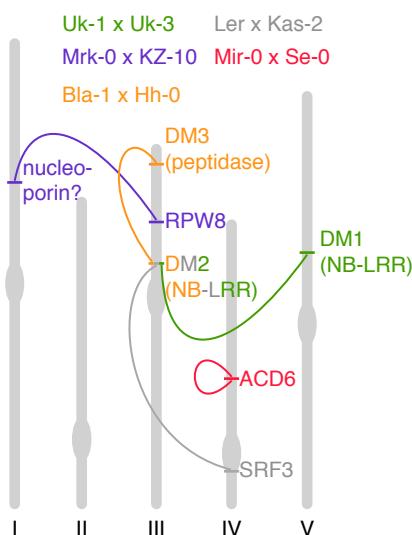


Fig. 4. Location of hybrid necrosis genes on the 5 chromosomes of *A. thaliana*. Mapping crosses indicated on top. The Ler x Kas-2 F₂ case has been described by others⁴⁰.

new variants are generated, not only through single site mutations, but also through gene conversion and unequal cross over. They share these properties with the *OAK* locus, at which interacting alleles cause growth abnormalities in yet another hybrid case that we have characterized⁴⁵. Together, this indicates that certain regions of the genome are particularly prone to spawning genetic incompatibilities.

is best interpreted as the immune system being primed for defense². Moreover, we have previously shown that about 20% of all *A. thaliana* accessions, both globally and in local populations, carry a special *ACD6* allele that confers broad-spectrum resistance to a wide range of pathogens. This allele greatly reduces the growth of inbred strains, similar to what we see in hybrid necrosis individuals, and we have proposed that *ACD6* mediates a **fitness trade-off between immunity and growth**⁴⁴. Of the two interacting hybrid-necrosis inducing *ACD6* alleles, one is old and broadly distributed, while the other is only found in the Northeast of Spain. Both alleles have risen to high frequency in this region, suggestive of **local adaptation**. We are excited that both alleles co-occur in local populations, and that hybrids can be found in the field (Fig. 5).

Common to most of these loci is that the **encoded proteins** have **internal repeats** (ankyrin repeats, coiled coil repeats, LRRs) and that the genes themselves occur in **tandem arrays**. These features likely increase the frequency with which

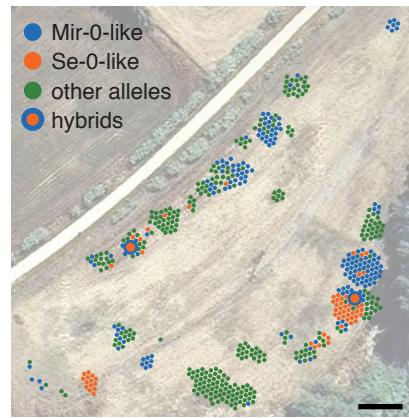


Fig. 5. Distribution of individuals carrying different *ACD6* alleles at the CB16 collection site near Llagostera, Spain. Size bar = 10 m.

Other cases of potentially deleterious epistatic interactions

We have never observed obvious hybrid sterility in *A. thaliana* crosses that produced flowers, except when parents differed in ploidy². However, there is another type of epistatic interaction that can be inferred from the **absence of certain genotypic combinations** at unlinked loci in F₂ or F₂-like populations. This can be due to inadvertent selection, e.g., because late-germinating lines are not propagated, but several cases are associated with lethality of specific segregants, including ones that we initially identified^{46,47}. Olivier Loudet has characterized three striking examples, in which the causal genes were **redundant paralogs** with reciprocal patterns of inactivation in different lineages⁴⁸⁻⁵⁰. Since such mutations were always found in several accessions, one might expect that it should be possible to detect long-distance linkage disequilibrium (LD) in *A. thaliana*. Paradoxically, we have not yet been able to do so. One possible explanation is that allelic diversity at the causal loci interferes with LD detection. This warrants further investigation.

The next step

Arabidopsis thaliana is an excellent model for investigating epistatic interactions with clear phenotypic consequences in both F₁ and F₂ generations. However, all cases studied so far were discovered in **random crosses**, and while the causal genes indicate tantalizing trends, we lack a good understanding of the types of genes involved in these interactions, and of the potential selection pressures underlying the evolution of the causal alleles. Moreover, we do not know how much of what we observe in *A. thaliana* is **due to the inbreeding mating system**. I am therefore proposing a **systematic study** of F₁ and F₂ interactions that **activate the immune system or affect normal development**. We will reveal subtle interactions by testing with molecular and imaging methods for activation of the immune system in F₁ hybrids, and by high-throughput ultra-dense genotyping for segregation distortion in F₂ progeny. Because the rate of selfing affects the fixation of alleles with deleterious epistatic interactions in populations, we will extend our studies to *A. lyrata*, a sister species with obligatory **outcrossing**. And to understand what fuels the evolution of epistatic interactions among components of the immune system, we will **link allelic diversity at immune genes to microbial diversity in the wild**. I believe that IMMUNEMESIS takes in an exemplary manner advantage of genomics-enabled genetics tools and resources. Given our leading role and proven success in developing such tools and resources, I am optimistic that our ambitious program, with its complementary and innovative strands, will be highly productive.

Section b. Methodology

Under our first aim, we will exploit a **unique resource—a complete diallel of 6,400 crosses** among 80 fully sequenced *A. thaliana* strains—for analyzing genome-wide epistatic interactions with easily detectable phenotypic effects. This will be complemented under the second aim by comparing patterns of immune system activity in **inbreds** and **hybrids** at the molecular level, and under the third aim by investigating more subtle interactions that are only expressed in the F₂ generation. Under the fourth aim we will compare the findings from *A. thaliana* to the **outcrossing species** *A. lyrata*. Finally, under the fifth aim, we will begin to reveal associations between **microbial communities** and **host genetics** in nature.

As a **prerequisite for the first three aims**, we are currently developing an **80x80 diallel**. We have recently published the first major phase of the *A. thaliana* 1001 Genomes project, reporting on 80 strains representing eight geographic regions from the native range of the species²¹ (Fig. 6). I am proposing to take advantage of these genome sequences for systematic forward genetic analyses of epistasis. To facilitate a large number of crosses, we have made all accessions male-sterile, by knocking down the *AP3* gene using artificial miRNAs (amiRNAs), a technology developed in my lab⁵¹. We chose *AP3*, because its loss of function leads to the conversion of stamens into carpels⁵², and we can easily monitor successful suppression of *AP3*. We have a potent amiR-*AP3* construct with which we can reliably produce male-sterile plants for all accessions. We have previously not observed evidence of parental effects in hybrid failure. However, reciprocal hybrids can differ in growth⁵³⁻⁵⁶. Additionally, silencing of resistance genes has been reported⁵⁷⁻⁶⁰, and there are parent-of-origin components to gene silencing⁶¹. We will therefore produce a complete diallel, in which every accession is crossed to every other in both directions, with at least 100 seeds per cross. The transgene will not confound subsequent studies, because we use primary, heterozygous transformants, and we can easily obtain non-transgenic F₁ plants. Since we initiated the crossing program last year, we have produced over half of all diallel crosses, and anticipate that this resource will be available at the start of IMMUNEMESIS.

Aim 1: A systematic study of dominant hybrid necrosis alleles in *A. thaliana*

Rationale: Although *NB-LRR* genes are among the most intensely studied genes in *A. thaliana*, there have been relatively few studies of the geographic patterns of *R* gene alleles^{29,62-66}. One of the reasons for the paucity of such studies is the extraordinary sequence diversity of *R* genes, which makes PCR-based analyses difficult. In addition, the effects of individual *R* gene alleles can be background-dependent⁶⁷. Based on the little that is known, we hypothesize that epistatic interactions that activate the immune system are not necessarily more common between geographically distant accessions than within the same geographic region. Moreover, the causal alleles in two of our cases co-occur locally, suggesting that such interactions could have ecological and evolutionary impacts in the natural environment.

Our knowledge of epistatic interactions comes mostly from quantitative genetics in random crosses, or from genome-wide expression analyses irrespective of organismal phenotypes⁶⁸. Our **80x80 diallel** will inform us in an unbiased manner about the **frequency and distribution of an important subset of epistatic interactions** that are **biologically meaningful**. Specifically, we will be able to determine how often common alleles that might be under balancing selection indicative of fitness trade-offs contribute to autoimmunity.

Hybrid necrosis studies: We will grow all 6,400 F₁ **hybrids** of the **80x80 diallel** in short days (to delay flowering) at 16°C. The 80 self-crosses among these will serve as controls. Hybrid necrosis is often temperature sensitive, a common property of immune responses⁶⁹, and known *A. thaliana* cases are strongly expressed at 16°C. For subsequent genetic studies, the symptoms can be suppressed by moving plants to higher temperatures that mimic abiotic stress, and seeds can be obtained. In addition to visual inspection for HR-like symptoms and dwarfism, we will use **fluorescence imaging** at 4 and 6 weeks of age to detect reduced chlorophyll content, an early indicator of necrosis⁷⁰. If we cannot clearly distinguish two phenotypic classes, we will focus on the bottom tail of 5% for chlorophyll content. Candidates for aberrant immune responses will be regrown and examined for **microscopic signs of cell death**, for increased expression, as measured by qRT-PCR, of **immune response markers** *PRI*, *LOX2*, *PDF1.2* and *FRK1*, which are induced downstream of pathogen attack⁷¹⁻⁷⁴, and for **growth** of the mildly virulent **model pathogens** *P. syringae* pv. *maculicola* ES4326⁷⁵ and *P. syringae* pv. *tomato* DC3000⁷⁶. To allow quantitative analysis of bacterial growth curves in many plants, we will infect plants with strains that have been tagged with a bacterial luciferase that does not require an exogenous substrate⁷⁷. Parents and random non-affected hybrids will serve as controls.

Previously, we found that about 2% of 861 random crosses resulted in hybrid necrosis². With 3,160 non-self, non-reciprocal crosses, we expect **about 60 macroscopically visible hybrid necrosis cases** and a more difficult to predict number of **milder cases**. We will identify causal genes by combining **mapping-by-**

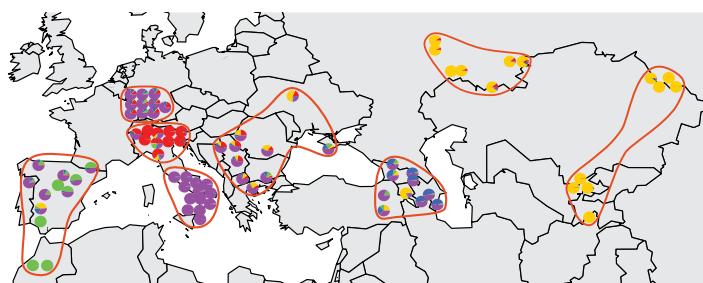


Fig. 6. Provenance of 80 accessions sequenced in the first major phase of the 1001 Genomes project²¹. The eight geographic regions sampled are indicated. Pie charts indicate STRUCTURE results, reflecting genome-wide relatedness and diversity.

of a two-gene, fully dominant interaction, we expect 1:3 ratios of parental alleles at the causal loci in normal individuals. A single region where no individual is heterozygous will indicate a single-locus interaction. (We also have experience with more complex situations.) The diallel design will allow us to identify accessions that likely have the same causal alleles, thus potentially reducing the number of crosses that need to be analyzed. We will exploit local haplotype sharing to pinpoint the causal loci. Together, these approaches will enable our ambitious program of mapping genes in dozens of crosses between accessions. Confirmation of having identified the causal genes will eventually come from knockdown and retransformation in *A. thaliana*, along with reconstitution of the interaction by transient expression in *N. benthamiana*, as we have done before. Once the causal genes have been identified, we can answer questions such as whether hybrid autoimmunity is similarly frequent in geographically close and distant accessions (since known resistance alleles seem to show little geographic structure^{29,62-66,81-84}), and whether specific regions of the genome are overrepresented for such interactions. And in addition to functional analyses of the encoded protein variants, we will of course analyze the population genetics of these genes.

Linking autoimmunity to patterns of *R* gene diversity: A potential challenge for final identification of the causal genes is that these may reside in polymorphic gene clusters. Such clusters are difficult to reconstruct from remapping of short reads to the reference genome. As a first step towards a better description of diversity in the *NB-LRR* family, the major type of *R* genes, we have exploited coverage profiles to classify accessions according to their alleles. These analyses revealed three broad classes of genes: one with predominant presence/absence polymorphisms, one with mostly intermediate levels of coverage, and a third group with little difference from the reference. These observations mirror prior, more limited studies (e.g., ref. 85). Our straightforward approach has confirmed patterns for two well-characterized *R* genes, *RPM1*⁶² and *RPS2*⁶⁴ (Fig. 7). To support the exact identification of causal genes in complex regions of the genome, we will first extend our analyses to include SNPs, small indels and targeted assemblies of left-over reads²¹. The gold standard will be complete assemblies across *NB-LRR* and other complex gene clusters. To this end, we have identified 30 accessions that capture much of the common diversity in the native range of *A. thaliana*. We will assemble their genomes from Illumina libraries with insert sizes of up to 40 kb. Using the Col-0 reference genome for benchmarking, we can assemble 99% of all *NB-LRR* genes with multiple libraries and high genome coverage. The 30 reference genomes will be integrated into a single genome graph so that reads from the 80 accessions can be aligned simultaneously against all genomes, following ideas we have laid out before⁸⁶. In addition, we may design capture arrays based on the assembled genomes, for targeted analysis of *NB-LRR* and other complex genes, a method that we have experience with^{87,88}.

The primary motivation for this work is to support the forward genetic studies in our program. Detailed profiling of *NB-LRR* genes and other complex regions of the genome will help us to identify the causal genes, and tell us which

sequencing techniques pioneered in my lab with RAD-seq analysis⁷⁸⁻⁸⁰. Our experience has been that sequencing of pooled individuals is not always the best approach when genetic architecture is unknown. For each F₂ population, we will barcode 192 normal individuals, and sequence these in half a lane of an Illumina HiSeq 2000 flow cell; with genome complexity reduction to 5%, this will produce 10x coverage at each marker. Our RAD-seq protocol generates about 2,500 markers, which is more than the total number of recombination events in 192 F₂ progeny⁴⁷. In the simplest case

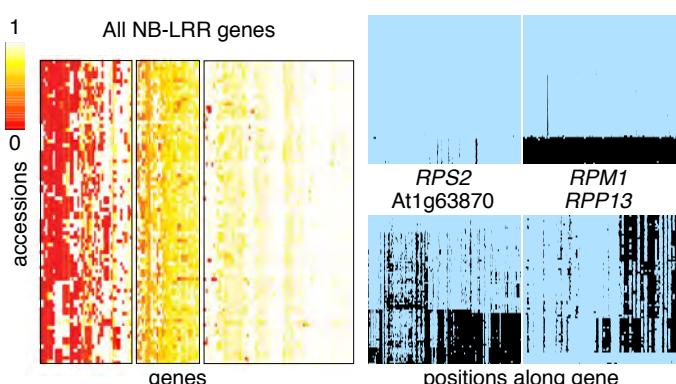


Fig. 7. Fraction of reference *NB-LRR* genes covered with short reads from 80 accessions. Left, Overview for all 145 *NB-LRR* genes; color bar indicates coverage of individual genes in each accession. The three broad patterns are discussed in more detail in the text. Right, examples of coverage of four genes along their length, with accessions clustered by coverage profiles. *RPM1* belongs to the left-hand class, *RPS2* to the right-hand class. At1g63870 and *RPP13* are examples of the middle class of genes with a complex pattern of diversity.

ones are especially prone to **spawning autoimmune interactions**. We are especially keen to know how exceptional the *DM2* region is, considering that three of five known hybrid necrosis cases involve this locus (Fig. 4). By including defined populations from relatively small geographical areas (Swabia and S. Tyrol²¹), we can begin to look at possible **local selection pressures** that affect the prevalence of causal loci. Once we know the exact alleles, we can go also back to the populations they come from and determine with greater precision the local distribution of these alleles and their possible interactions with microbes (see Aim 5).

Our efforts will also reveal the first, exquisitely **detailed picture of genome- and species-wide diversity of *NB-LRR* and other *R* genes**. Of particular interest will be the question of geographic patterns. For example, we know that marginal populations have reduced genetic diversity, but does this also hold for immune loci? In our unpublished work with *Capsella rubella*, a young species that recently went through an extreme genetic bottleneck^{23,24}, we have found hints that diversity around *R* gene cluster far exceeds the genome average. If generally true, the question arises how this is achieved, as it would imply that genomic regions containing *R* genes are **privileged** when it comes to **exchange of alleles** between populations.

Contingency planning: *We have established the mapping and genotyping methods. We cannot predict how many new hybrid necrosis cases we will find, nor how many weaker autoimmune interactions we can detect. However, there is no reason to believe that the 80 accessions are unusually devoid of hybrid necrosis genes. The most difficult aspect will be the unambiguous identification and assembly of causal genes. If this turns out to be more challenging than anticipated, we will resort to Illumina sequencing of individual fosmids⁸⁴, although this will reduce the number of cases we can investigate in detail.*

Aim 2: The spectrum of subtle epistatic interactions priming the *A. thaliana* immune system

Rationale: As discussed above, epistatic interactions that trigger autoimmunity and at the same time reduce biomass are not that rare. On the other hand, outcrossing of inbred strains often has positive effects on plant performance, a phenomenon known as heterosis, and *A. thaliana* is no exception⁵³⁻⁵⁶. Remarkably, the literature has little to say about the **relationship between vigor of growth and disease resistance in hybrids**. This begs the question how the connection between the immune system and growth is managed in hybrids. There are more opportunities for epistatic interactions between non-co-evolved alleles in hybrids than in inbred strains, so our expectation is that the **background level of immune system activity** is on average higher in hybrids. We will test this hypothesis.

A major part of Aim 2 includes GWAS approaches; the reason that we propose a separate aim for these is that the 80x80 diallel from Aim 1 is highly structured (because we want to understand geographic patterns). In addition, we will determine growth trajectories, measure expression of immune system marker genes and assay growth of pathogens in all 1,440 genotypes of Aim 2. These analyses would be difficult to perform for the entire set of 6,400 genotypes in Aim 1, but we will include genotypes that connect the two aims.

Comparing inbreds and hybrids: To determine the extent of epistatic interactions affecting immunity and biomass, we will compare three sets of plants, totaling 1,440 genotypes. The first set will include 440 diverse Eurasian inbred **accessions** that have been genotyped at high density¹⁷ and that are being resequenced in the 1001 Genomes project, plus 32 parents of our 80x80 diallel, representing three geographic regions, N. Africa/Iberian Peninsula, Swabia and S. Tyrol²¹. This number of genotypes, when appropriately chosen, provides substantial power for GWAS in *A. thaliana*^{16,17,44,89,90}. The second set will consist of 472 **F₁ hybrids** with the first group of 440 lines as parents. We will preferentially choose parental combinations with a high number of differences at intermediate-frequency alleles, to maximize power of single-locus GWAS and possibilities for epistatic interactions. The third set will be a **subset of the 80x80 diallel**, namely a half diallel of 496 F₁ hybrids from the 32 accessions mentioned above; together, they will connect the genotypes from Aim 2 and Aim 1. All genotypes will be grown in short days at 16°C, 20°C and 23°C, because immune responses are often temperature sensitive (e.g., ref. 69). In addition to imaging with visible light at weekly intervals for total **plant size**, we will use **fluorescence imaging** to detect reductions in chlorophyll content, an early indicator of the onset of necrosis⁷⁰. After six weeks, plants will be harvested, dried and **weighed**. From a separate set of plants grown for two weeks, we will determine in triplicate samples the **expression** of four **immune response markers**, *PRI*, *LOX2*, *PDF1.2* and *FRK*⁷¹⁻⁷⁴. Finally, we will use luciferase-based imaging to measure **growth** of the mildly virulent pathogens *P. syringae* pv. *maculicola* ES4326 and pv. *tomato* DC3000⁷⁷.

We will analyze traits individually and jointly, after appropriate transformation of values. Because transgressive variation is common in *A. thaliana*⁹¹⁻⁹⁴, we expect the distribution of trait values in the F₁ hybrids to be broader than in the inbred strains. The more interesting question is whether there are **systematic**

differences between hybrids and inbreds. Specifically, if the immune system is often “on edge” and poised to detect proteins it has not encountered before, we may find not only increased variance of immune system activity in hybrids, but also a **higher mean** than in inbreds. To address this hypothesis, we will first ask how often individual hybrids exceed mid-parent values. In addition, we will fit a regression model for predicting the measured traits based on all SNPs in the inbreds, and apply this model to the hybrids. If our hypothesis that the hybrids contain a different spectrum of epistatic interactions than the inbreds is correct, then the **model from the inbreds will not generalize to the hybrids**.

We will ask whether **genome-wide divergence** between any two parents is predictive for immune system activity in hybrids. We can pose the same question for ***R* genes only**, by exploiting knowledge about complex *R* gene loci generated under Aim 1. Furthermore, by including the half diallel from N. Africa/Iberian Peninsula, Swabia and S. Tyrol, we can test whether accessions from a genetically more uniform area such as S. Tyrol behave differently from those that come from genetically more diverse areas. Finally, this material will allow us to ascertain whether there is any **systematic correlation between a primed immune system and biomass**. While mutants in which the immune system is hyperactive are often small¹, it is unclear whether the same response exists in wild-type material. Obviously, being able to sever the trade-off between growth and immunity would have enormous implications for plant breeding.

GWAS: To understand genome-wide patterns of epistasis, we will perform comparative **GWAS** for immunity traits and biomass in inbreds and F₁ hybrids. We are in a privileged position because Karsten Borgwardt is a colleague on campus with whom we are extensively collaborating (e.g., ref. 21,95). His group is not only developing very fast algorithms for conventional GWAS that are neither limited by the number of markers nor the number of individuals (FaST-LMM, which greatly speeds up the popular EMMA algorithm without relying on heuristic assumptions as in EMMAX^{96,97}), but also implementing innovative approaches for detection of epistatic interactions⁹⁸⁻¹⁰⁰. We will compare the results of single-locus GWAS in inbreds and hybrids (extending GWAS to include heterozygotes is trivial, as this is the norm in most species). If we find loci that only appear in hybrids, or that produce a stronger signal in hybrids, this will be a first indication of possible epistatic interactions. We will then search the genome by GWAS for possible **epistatic interaction partners** for these candidates. In parallel, we will perform GWAS for epistatic interactions with polymorphic immune genes, including the ~200 *NB-LRR* loci and other *R* genes such as *RPW8*⁴³ and *MLO*¹⁰¹ as well as *PRR* loci¹⁰², and compare these to the background of epistatic interactions among all other genes. We hypothesize that the behavior of hybrid necrosis loci is indicative of a network of weaker epistatic interactions that prime the immune system. If this is correct, then **immune genes** should become **enriched** in the epistatic interaction scans. Similarly, we will ask how often the same loci affect immune system activity and biomass. Finally, while the primary motivation for this aim is not to find candidates for experimental validation, but rather the testing of general hypotheses about immune system behavior, we will of course not ignore immune system loci discovered in this part of IMMUNEMESIS.

Contingency planning: As with any GWAS experiment, we cannot predict whether or how many significant associations we will detect. Note, however, that we aim to compare the behavior of entire groups of *a priori* defined genes, and that the usual multiple hypothesis testing problems do not interfere with our goals. The power of GWAS can normally be increased by adding genotypes, which we may consider. In summary, even if we identify few individual loci responsible for differences in priming of the immune system, we will generate valuable knowledge about the sum of epistatic interactions that affect the immune system, and how these contribute to potential trade-offs with vegetative growth.

Aim 3: A systematic map of deleterious recessive epistatic interactions in *A. thaliana*

Rationale: Several epistatic interactions with strong morphological phenotypes manifest themselves only in the F₂ generation of *A. thaliana* crosses^{41,47-50} (Fig. 8). However, given the quantitative effects of a large fraction of natural alleles, it is not unlikely that many such interactions already reduce plant performance in the F₁. We therefore propose to systematically compare the epistatic F₁ interactions discovered in Aim 1 with potentially deleterious F₂ epistatic interactions. This will tell us, for example, whether recessive epistatic interactions show a bias towards specific genes, as with the F₁ cases. To this end, we will analyze segregation distortion in all 6,320 non-self F₂ populations derived from the 80x80 diallel.

Extending systematic mapping of deleterious epistatic interactions to the F₂ generation: We will harvest F₂ seeds from the (non-transgenic) F₁ plants studied under Aim 1. We will sow at least 200 F₂ individuals from each cross onto Petri dishes. To reduce the effects of differences in dormancy, we will use an after-ripening period of at least six months; in addition, seeds will be stratified for six days in the dark, and germination behavior of parents will be taken into account in subsequent analyses.

After we have ascertained that germination has been reasonably uniform, we will look for any morphologically **abnormal** plants that segregate in each population after two weeks of growth. We expect these to appear at a frequency of at most 1:7 (if a recessive locus from one parent interacts with a dominant locus from the other parent), and more likely at 1:15 (if two recessive loci interact) or lower. In such cases we will grow out larger numbers of individuals, before genotyping individuals as described under Aim 1. For **phenotypically normal populations**, we will process all plants as a **pool** for RAD-seq analysis⁸⁰, combining 48 to 96 populations in each lane of an Illumina HiSeq 2000 flow cell for about 16 to 32€ consumables per cross. With our protocol, we will generate greater than 40-80x coverage at each of about 2,500 markers, spaced on average 50 kb apart. To compare genome-wide allele frequencies across populations, a beta-binomial based summary statistic will be used to **model allele frequency** and its **variance** along the genome. Regions with frequencies that significantly deviate (established by permutation tests) from the expected **1:1 segregation ratio** of parental alleles will be identified for each population and their locations compared between populations. GWAS (for details, see Aim 2) will be employed for simultaneous analysis of all crosses, including those that share a single parent. Based on simulations for different coverages, we expect to detect any case of **segregation distortion** that reduces one parental allele by as little as 12-18%. Thus, we will be able to not only identify **lethal**, but also **sub-lethal** interactions.

We will combine information from crosses in which the same regions of the genome show parental bias. Fine mapping will be further supported by performing local association mapping in these regions. Two caveats are the possibility of extragenic suppressors, or that the interacting alleles only delay germination, but do not kill affected individuals. To take these scenarios into account, we will make use of the **1001 Genomes** information (<http://1001genomes.org>); if interacting alleles have a non-suppressible lethal effect, they should not co-occur in natural accessions. A subset of crosses, representative of the different cases identified, will be studied in more detail; we will determine at which stage affected genotypes die, and confirm causal genes with knockdowns using amiRNAs⁵¹ and transformation with genomic constructs.

Together, this work will reveal the **spectrum of genes** that can **interact to reduce** the frequency of specific **genotypic classes** in F₂ populations. Questions that we will be able to answer include whether such genes are biased towards certain **functional categories**—and if so, whether **immune loci** are among them—and whether they share specific structural features, as suggested for some of the known F₂ cases⁴⁸⁻⁵⁰.

Contingency planning: Based on the literature¹⁴, it is almost certain that we will detect many examples of segregation distortion. Because there can be allelic heterogeneity at the causal loci^{48,49}, local association studies might not be very powerful. We will then have to rely on mapping in F₂ populations alone, for which we will need to increase mapping resolution, by analyzing more plants and more markers. This will reduce the total number of cases we can study in detail, but it will not pose a principal hurdle.

Aim 4: Epistatic interactions in inbreeding versus outcrossing *Arabidopsis*

Rationale: In a predominantly inbreeding species such as *A. thaliana*, slightly deleterious mutations accumulate more easily than in outcrossers¹⁰⁴. Mutations that are only deleterious in hybrids are also likely to be more prevalent. On the other hand, mutations with strongly deleterious effects when homozygous should be purged more effectively. It is thus not obvious how common hybrid necrosis and other forms of potentially deleterious epistatic interactions will be in an outcrosser compared to a selfer. The genus *Arabidopsis* is an excellent choice for addressing this question, as we have a high-quality genome sequence of the outcrossing species *A. lyrata*²². I propose to determine how frequent **hybrid necrosis** is in *A. lyrata*, and to search for evidence of long-distance LD indicative of **incompatible regions** in the genome.

Prevalence of hybrid necrosis in *A. lyrata*: Because European *A. lyrata* populations are generally self-incompatible¹⁰⁵, it is more cumbersome to maintain wild genotypes in the lab and to perform controlled crosses among them than it is for *A. thaliana*. However, the obligate outcrossing behavior also offers a **unique opportunity**, because in contrast to *A. thaliana*, any seed collected from wild *A. lyrata* individuals will be the product of a **natural cross**. Through a collaboration with Markus Koch (Director of the Heidelberg Botanical Garden), we have access to several populations in Germany, Austria and the Czech

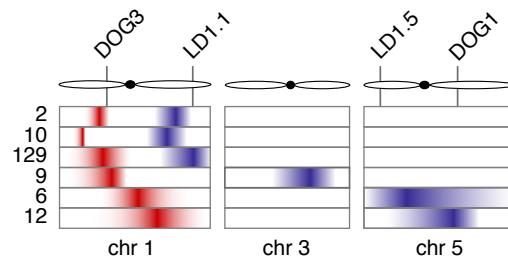


Fig. 8. Six maps of F₂ segregation distortion involving two loci⁴⁷. The locations of two *DELAY OF GERMINATION* (*DOG*¹⁰³) loci and two known interacting loci on chromosomes 1 and 5 (*LD1.1* and *LD1.5*, ref. 48) are indicated on top. Color intensity reflects p-values for deviation from expected allele frequency, with darker for lower p-values.

Republic. We will collect seeds from at least 25 individuals from eight sites. We will sow 10 seeds per family and monitor seedlings for signs of (**hybrid**) **necrosis** over at least two months of growth at 16°C short days. Potential candidates that express strong necrosis will be transferred to 23°C, to rescue the plants and allow them to flower². Pollen from these candidates will be crossed onto the MN47 accession, from which the reference genome sequence was produced²². Because MN47 is self-compatible, it can be used as crossing partner regardless of the self-incompatibility alleles present in the other individual. If necrosis in the tester individual was due to two dominantly acting loci, we expect one quarter of the F₁ to express necrosis again. We will backcross these individuals to MN47 and **map introgressed loci** by RAD-seq analysis of at least 100 BC₂ individuals. *Arabidopsis lyrata* populations from Central Europe harbor much more diversity than found over much greater geographic distances in *A. thaliana*^{106,107}. Thus, we expect to have many markers that can be used for genetic mapping. While 100 BC₂ individuals provide only limited mapping resolution, we hypothesize that hybrid necrosis loci in *A. lyrata*, if they can be found, will also encode immune proteins, facilitating the identification of candidate genes. Although genetic transformation of *A. lyrata* is more laborious than that of *A. thaliana*, it is possible¹⁰⁸, and we will attempt to **validate candidate genes** both by amiRNA-mediated knockdown⁵¹ and by reconstitution in the heterologous *N. benthamiana* system.

Detection of deleterious epistatic interactions from resequencing of local *A. lyrata* populations: Because *A. lyrata* is relatively rare, there should be little gene flow between stands, and each stand should behave like a **randomly mating population**. This offers opportunities for scanning the genome for **deleterious epistatic interactions** by identifying physically unlinked regions that are in LD. To this end, we will investigate **two large populations**, both of which have several thousand individuals and are not endangered, one at Spiterstulen in Norway and the other near Plech in Germany (M. Koch, pers. communication). We will first sequence a pilot set of 30 individuals from each stand at 20x genome coverage, to determine the extent of local LD in each population. The results will inform us about the average length of haplotype blocks, and they will guide the approaches for analyzing at least 500 individuals from each population. (If haplotype blocks are long, we can use RAD-seq instead of whole-genome sequencing.) We will apply Hidden Markov Models developed for crosses^{109,110} to segment the genome and assign ancestry in each haplotype block. This information will be used to identify pairs of haplotype blocks that co-occur much less than expected by chance (ideally, never) in the individuals of each population. We will consider both dominant and recessive interactions. If we find such cases, we will study these further by **crossing individuals** that segregate for the alleles involved, and genotyping their offspring. The exact crossing design will depend on whether the interactions appear to be dominant or recessive, and whether we can identify individuals that are homozygous for the candidate loci. We aim to perform the genotyping while the plants are still growing (*A. lyrata* individuals can be easily maintained for months or even years), so that we can identify the corresponding genotypes among the sequenced material.

Contingency planning: Given that 2% of crosses in *A. thaliana* produce hybrid necrosis, that we have found two cases involving accessions from the same local populations² (Fig. 5) and that diversity in local *A. lyrata* populations is much higher than in *A. thaliana*^{106,107}, failure to find hybrid necrosis in *A. lyrata* would be an important result. Similarly, not finding evidence for recessive deleterious interactions would allow important inferences about the strength of selection against such interactions in outcrossing *A. lyrata*.

Aim 5: Natural microbial communities driving immune system diversity in *A. thaliana*

Rationale: A possible criticism of IMMUNEMESIS is that assays with pathogens play only a minor role in our work. Indeed, we often take the expression of immune genes and morphological symptoms in unchallenged plants as a proxy for their ability to defend themselves against pathogens. Several lines of evidence support this assumption. Plants that express autoimmunity are often more resistant to pathogens⁵⁻⁹, and such mutants are even employed in breeding for resistance^{10,11}. Furthermore, it is difficult to generalize from laboratory assays of immune responses to the field, given that the **distribution of pathogen genotypes** on natural populations of even well-studied plants such as *A. thaliana* is largely unknown.

Three major natural pathogens have been described in the native range of *A. thaliana*: powdery mildew *Golovinomyces orontii*, downy mildew *Hyaloperonospora arabidopsis*, and white blister rust *Albugo laibachii*¹¹¹⁻¹¹³. Reference genome sequences are available for all three species¹¹⁴⁻¹¹⁶. Among the three pathogens, resistance to *H. arabidopsis* has been most intensively studied, because of the interesting genetic complexity on both the pathogen and the host side¹¹⁷. A large fraction of *R* genes cloned in *A. thaliana* confer resistance to *H. arabidopsis*. They are called *RPP* genes, for *RECOGNITION OF PERONOSPORA PARASITICA*, and our most common hybrid necrosis locus, *DM2*, is an *RPP1* homolog. *RPP* genes known to confer resistance of *H. arabidopsis* races are located throughout the *A. thaliana*

genomes, with some having alleles that mediate recognition of other pathogens. Of particular interest is *RPP13*. It has many different alleles, and the encoded proteins interact with different versions of the ATR13 effector injected by different *H. arabidopsis* strains into the host¹¹⁸. Several authors have studied patterns of interactions between different *H. arabidopsis* isolates and *A. thaliana* strains⁸¹⁻⁸⁴; one fact gleaned from these studies was that there was no obvious geographic structure in the distribution of resistances among *A. thaliana* strains. Different from resistance to *H. arabidopsis*, which occurs at intermediate frequency, resistance to *A. laibachii* is rare^{113,119,120}, while resistance to *G. orontii* is common¹²¹.

Remarkably, despite the impressive progress made with laboratory studies of these pathogens, there are no systematic reports of their distribution in natural populations, nor how their distribution maps onto genetic diversity in the host. Moreover, previous work on the microbiome of natural *A. thaliana* populations has targeted only specific bacterial species, in the case of foliar microbes¹²²⁻¹²⁸, or explored only controlled conditions, as in the case of the root microbiome^{129,130}. As an alternative we therefore propose an ambitious, two-pronged program. First, we will relate **distribution and genetic diversity** among the three **filamentous pathogens** *H. arabidopsis*, *A. laibachii* and *G. orontii* to the genetic diversity of **local *A. thaliana* populations**. This is directly pertinent to the other aims of our proposal, since of the known hybrid necrosis genes, one, *DM2*, is a homolog of *RPP1*. Another one is an allele of *RPW8*, which mediates resistance to *G. orontii*⁴³, and *ACD6*, which we have also found to be involved in hybrid necrosis, can quantitatively increase resistance to *G. orontii*⁴⁴. We will complement these targeted studies with **an unbiased and sensitive description of the microbiome** associated with the above-ground part of naturally grown *A. thaliana* plants. Together, this will establish the **links that exist between the microbiome and genetic diversity of the host**.

Sampling natural populations: Our medium-term goal is to analyze local *A. thaliana* populations in which causal genes for epistatic interaction that activate the immune system have been identified (Aims 1-3). Initially, we will focus on local populations around Tübingen that contain several genotypes and that we have been following for several years¹³¹. By RAD-seq we will first establish that there are **large patches of uniform genotypes**, as we have found before in several populations¹³¹. In the next year, sampling of populations will start when rosettes are 1 cm in diameter, following natural germination in spring (February/March) and fall (August/September). Sampling will be repeated in monthly intervals until the end of the life cycle. We will record the locations of individuals within each site, and collect leaf punches from each plant for RAD-seq genotyping.

We will classify plants according to the macroscopic presence or absence of pathogens, and presence or absence of disease symptoms such as HR. Around Tübingen, normally a third to half of all plants are visibly infected by one of the **three filamentous pathogens**. We will mechanically retrieve the epiphytes from the leaf surfaces of such plants and process them for DNA extraction. Similarly, we will remove **microscopic epiphytes** from non-obviously infected leaves by washing with surface-active agents and by sonication. Commercial kits for DNA preparation from soil samples will be used on the epiphyte preparations and on the remaining leaf material, which will contain **endophytes**. DNA extraction will be repeated several times, to recover genomic DNA of microbes as completely as possible. Illumina libraries will be constructed with or without linear amplification by PCR, depending on the amount of DNA recovered.

The preparations of **mildew** or **rust** from obvious infections will be subjected to light-coverage shotgun sequencing or RAD-seq. We will ascertain the diversity of these samples by utilizing existing reference genomes¹¹⁴⁻¹¹⁶. Depending on the outcome of the initial experiments, we will adjust our strategies, such as pooling reads from samples that seem to contain the same genotype(s), for deeper analysis of variable regions of the genome, especially those encoding effectors. We will then test for associations between presence/absence of pathogens, pathogen genotype and host genotype, using spatial and temporal information and visual classifications as covariates.

For **unbiased analyses** of the microscopic epiphytic and the endophytic preparations, we will initially assess the suitability of both **16S rDNA amplicon** and **whole genome shotgun (WGS) sequencing**. (We will develop and establish methods for this aim during the first two years of IMMUNEMESIS. It is likely that improved sequencing methods and platforms will greatly facilitate large-scale WGS analyses.) For 16S rDNA analyses, we will use the Illumina MiSeq because of longer reads. We will amplify hypervariable 16S rDNA regions with different combinations of universal primers¹³², and align reads against multiple sequence alignments of known 16S rDNA sequences^{133,134}. MEGAN¹³⁵, written by my local long-term collaborator Daniel Huson at the University of Tübingen, will be used to identify operational taxonomic units (OTUs) and OTU frequencies will be determined from read counts. (We will also seek advice from my long-term collaborator Jeffery Dangl at UNC Chapel Hill, who recently published one of the first comprehensive analyses of the root microbiome of *A. thaliana* plants grown in controlled conditions¹³⁰.)

The preferred approach will be community profiling by **WGS sequencing**, since many effectors that determine host specificity are variable within pathogen species and one can gain only limited information from species-level classification. Note that we have found several percent of *G. orontii* reads in Illumina libraries generated from apparently healthy plants growing in the lab. Exploratory sequencing at high coverage (tens of Gb per sample) will indicate the minimum coverage required for the saturation of all microbial clades on individual plants. We will also test whether **de-enrichment of host DNA** by subtractive hybridization¹³⁶ improves the recovery of microbial sequences. After filtering out *A. thaliana* reads, the remaining reads from all samples will be combined for *de novo* assembly (with, e.g., MetaVelvet¹³⁷, MetaIDBA¹³⁸, SOAPdenovo¹³⁹). These assemblies will be added to reference datasets available in genome databases. For analysis of individual samples, reads will either be first assembled into longer sequences or aligned directly to reference datasets for **species identification and quantification**. Taxonomic assignment will result in phylogenetic trees from which **within (α) and between sample diversity (β)** will be assessed by computing distances with Unifrac and performing multidimensional cluster and principal component analyses^{140,141}, and by identification of metagenomic linkage groups¹⁴². We will also assess the suitability of machine-learning based classification in collaboration with my local Max Planck colleague and collaborator Bernhard Schölkopf. In addition, we will attempt to further improve our reference dataset by culturing the most abundant species and generating whole genome assemblies from these.

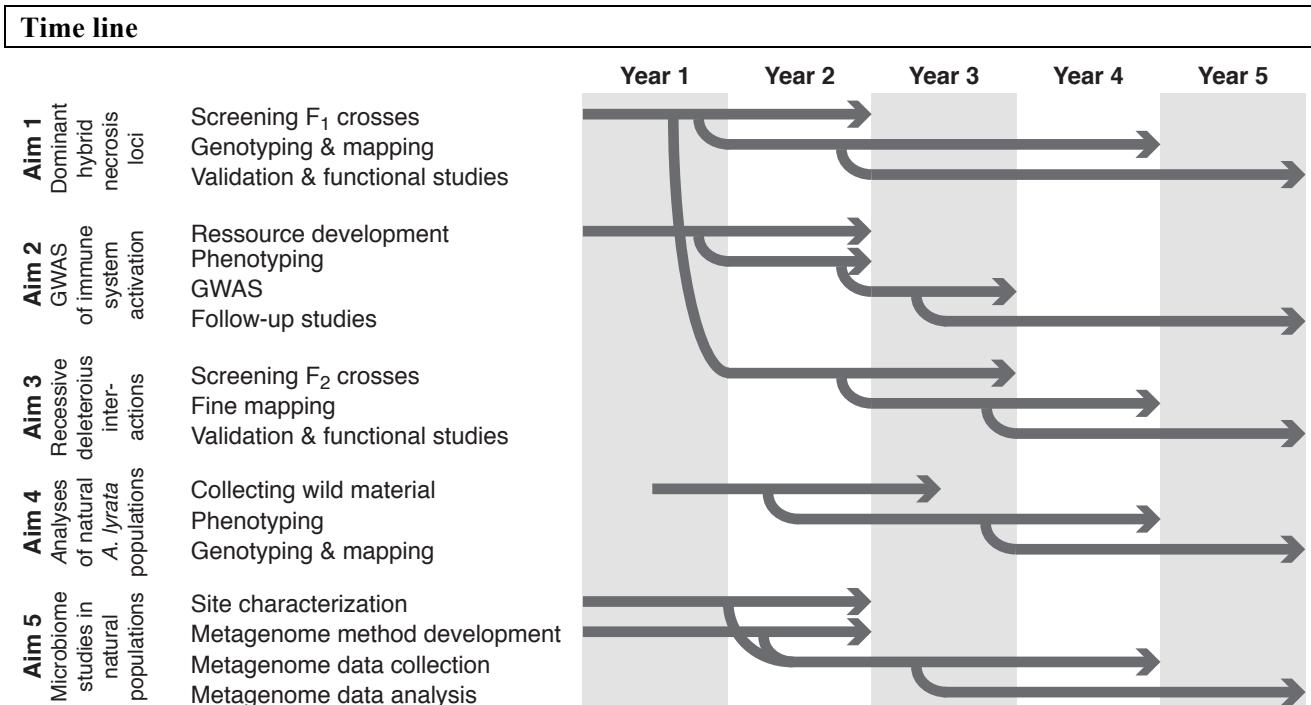
For linking microbial communities and host genotype, we will follow the general approach laid out in a recent study for human gut microbiota¹⁴². We will first determine whether there is more **variance between sampling sites than between host genotypes**, since it has been reported that epiphytic *Methylobacterium* community composition¹²⁸ is affected by collection site. We will then test whether different community clusters, determined as described above, differ significantly between *A. thaliana* genotypes at each site. A particular opportunity for further host genetics is offered by the fact that local populations can include **natural recombinant inbred lines** derived by intercrossing of a small number of founders, with large, easily identifiable haplotype blocks¹³¹. If we find evidence that host genotype affects the leaf microbiome, we can exploit such lines to determine whether **pathogen association and microbiome composition** can in principle be **mapped**, especially at sites where alleles that activate the immune system are found. Together, these efforts will be a crucial step in identifying actual causal agents that drive differentiation of the plant immune system in the wild.

Contingency planning: It has been shown before that similar *Methylobacterium* species can be detected on *A. thaliana* plants in subsequent years¹²⁸, indicating that microbial communities are not random. Research that attempts to link microbial communities and host genotype is a very active area, and it is very likely that additional, more powerful methods will be available by the time we generate our data. We will probably need to investigate large numbers (hundreds) of samples to establish statistically significant associations between microbial communities and host genotypes. For humans, such associations were obtained using less than 3 Gb of raw data per individual (and without replication of the host genotype)¹⁴². Given that a single Illumina HiSeq 2000 flowcell produces over 300 Gb, data generation should not be a limitation. In addition, sequencing technologies are likely to improve in a similar manner as they have over the past five years. In summary, while we cannot predict how easy or difficult it will be to find robust associations between microbial communities and host genotypes, the sort of experiments described here are absolutely essential if we are to understand what drives immune gene diversity in the wild.

Risk assessment and outlook:

ERC applications should be characterized by a high-risk/high-gain profile. While IMMUNEMESIS builds on the **expertise, resources and tools** my team has assembled over the past decade, it goes **beyond the current state of the art** by taking **full advantage of the very rapid developments** in the area of genomics-enabled genetics. I am not aware of any other program with multicellular organisms (except humans) that is similarly ambitious in performing large-scale forward genetic studies with natural material. Similarly, while laboratory studies of the interaction of *A. thaliana* with either natural pathogens or adapted pathogens from other species have been very successful in informing us about general principles and mechanisms of pathogen recognition, we are still largely in the dark when it comes to understanding patterns of co-evolution in natural populations. We will redress this situation in IMMUNEMESIS as well. Our aims span **a range of approaches with increasing risk**. Aim 1 will deliver rich knowledge about the network of genes that strongly activate the immune system. Aims 2 and 3 will extend this to more subtle interactions that are more difficult to detect. Under Aim 4, we will extend our work to *A. lyrata*; similar forward genetic studies have not been carried out before with wild, outcrossing species. Importantly, both positive and negative results

will be informative. Finally, I am very excited about the most ambitious part of the proposal, Aim 5. If successful, the systematic understanding of forces that shape the distribution of immune gene diversity in the wild will have important implications for engineering disease resistance in crops, by helping to chose the best ensembles of immune genes that maximize resistance without compromising heterosis of growth. In summary, I feel that IMMUNEMESIS contains the right mix of **straightforward and more uncertain elements**.



1. Robert-Seilaniantz A *et al*, *Annu Rev Phytopathol* 49:317 (2011).
2. Bomblies K *et al*, *PLoS Biol* 5:e236 (2007).
3. Bomblies K, Weigel D, *Nat Rev Genet* 8:382 (2007).
4. Krüger J *et al*, *Science* 296:744 (2002).
5. Walbot V *et al*, in *Genetic Engineering of Plants*, T. Kosuge, C. P. Meredith, A. Hollaender, Eds. (Plenum Press, New York, 1983), pp. 431-442.
6. Greenberg JT *et al*, *Cell* 77:551 (1994).
7. Jones JD, *Curr Biol* 4:749 (1994).
8. Dangl JL *et al*, *Plant Cell* 8:1793 (1996).
9. Rate DN *et al*, *Plant Cell* 11:1695 (1999).
10. Jorgensen JH, *Euphytica* 63:141 (1992).
11. Kamlofski CA, Acevedo A, *Agricult Sci* 1:56 (2010).
12. Allard RW, *Principles of Plant Breeding*. (John Wiley & Sons 2nd ed, 1999).
13. Wang Y *et al*, *Trends Plant Sci* 10:610 (2005).
14. Weigel D, *Plant Physiol* 158:2 (2012).
15. Clark RM *et al*, *Science* 317:338 (2007).
16. Atwell S *et al*, *Nature* 465:627 (2010).
17. Horton M *et al*, *Nat Genet* 44:212 (2012).
18. Ossowski S *et al*, *Genome Res* 18:2024 (2008).
19. Nordborg M, Weigel D, *Nature* 456:720 (2008).
20. Weigel D, Mott R, *Genome Biol* 10:107 (2009).
21. Cao J *et al*, *Nat Genet* 43:956 (2011).
22. Hu TT *et al*, *Nat Genet* 43:476 (2011).
23. Guo YL *et al*, *PNAS* 106:5246 (2009).
24. Foxe JP *et al*, *PNAS* 106:5241 (2009).
25. Boller T, Felix G, *Annu Rev Plant Biol* 60:379 (2009).
26. Tsuda K, Katagiri F, *Curr Opin Plant Biol* 13:459 (2010).
27. Jones JD, Dangl JL, *Nature* 444:323 (2006).
28. Thomma BP *et al*, *Plant Cell* 23:4 (2011).
29. Bakker EG *et al*, *Plant Cell* 18:1803 (2006).
30. Chisholm ST *et al*, *Cell* 124:803 (2006).
31. Dobzhansky T, *Genetics and the Origin of Species*. (Columbia University Press, 1937).
32. Muller HJ, *Biol Symp* 6:71 (1942).
33. Presgraves DC, *Nat Rev Genet* 11:175 (2010).
34. Rieseberg LH, Blackman BK, *Ann Bot* 106:439 (2010).
35. Touzet P, Budar F, *Trends Plant Sci* 9:568 (2004).
36. Jeukens MJ *et al*, *Plant Cell* 21:3368 (2009).
37. Yamamoto E *et al*, *Mol Genet Genomics* 283:305 (2010).
38. Rooney HC *et al*, *Science* 308:1783 (2005).
39. Shindo T *et al*, *PLoS ONE* 7:e29317 (2012).
40. Alcázar R *et al*, *Nat Genet* 42:1135 (2010).
41. Alcázar R *et al*, *PNAS* 106:334 (2009).
42. Tahir J *et al*, *Plant J* (2012).
43. Xiao S *et al*, *Science* 291:118 (2001).
44. Todesco M *et al*, *Nature* 465:632 (2010).
45. Smith LM *et al*, *PLoS Genet* 7:e1002164 (2011).
46. Balasubramanian S *et al*, *PLoS ONE* 4:e4318 (2009).
47. Salomé PA *et al*, *Heredity* 108:447 (2012).
48. Bikard D *et al*, *Science* 323:623 (2009).
49. Vlad D *et al*, *PLoS Genet* 6:e1000945 (2010).
50. Durand S *et al*, *Curr Biol* 22:326 (2012).
51. Schwab R *et al*, *Plant Cell* 18:1121 (2006).
52. Jack T *et al*, *Cell* 68:683 (1992).
53. Mitchell-Olds T, *Genetics* 140:1105 (1995).
54. Barth S *et al*, *Heredity* 91:36 (2003).
55. Meyer RC *et al*, *Plant Physiol* 134:1813 (2004).
56. Hochholdinger F, Hoecker N, *Trends Plant Sci* 12:427 (2007).
57. Stokes TL *et al*, *Genes Dev* 16:171 (2002).
58. Akimoto K *et al*, *Ann Bot* 100:205 (2007).
59. Zhai J *et al*, *Genes Dev* 25:2540 (2011).
60. Shivaprasad PV *et al*, *Plant Cell* 24:859 (2012).
61. Calarco JP *et al*, *Cell* 151:194 (2012).
62. Stahl EA *et al*, *Nature* 400:667 (1999).
63. Caicedo AL *et al*, *PNAS* 96:302 (1999).

64. Mauricio R *et al*, *Genetics* 163:735 (2003).
65. Rose LE *et al*, *Genetics* 166:1517 (2004).
66. Bakker EG *et al*, *Mol Ecol* 15:1405 (2006).
67. Van Poecke RM *et al*, *Plant Cell* 19:4046 (2007).
68. Mackay TF *et al*, *Nat Rev Genet* 10:565 (2009).
69. Whitham S *et al*, *Cell* 78:1101 (1994).
70. Chaerle L, Van Der Straeten D, *Trends Plant Sci* 5:495 (2000).
71. Ward ER *et al*, *Plant Cell* 3:1085 (1991).
72. Bell E *et al*, *PNAS* 92:8675 (1995).
73. Penninckx IA *et al*, *Plant Cell* 10:2103 (1998).
74. Asai T *et al*, *Nature* 415:977 (2002).
75. Dong X *et al*, *Plant Cell* 3:61 (1991).
76. Whalen MC *et al*, *Plant Cell* 3:49 (1991).
77. Fan J *et al*, *Plant J* 53:393 (2008).
78. Schneeberger K *et al*, *Nat Methods* 6:550 (2009).
79. Laitinen RA *et al*, *Plant Physiol* 153:652 (2010).
80. Baird NA *et al*, *PLoS ONE* 3:e3376 (2008).
81. Hall SA *et al*, *Mol Plant Pathol* 10:449 (2009).
82. Nemri A *et al*, *PNAS* 107:10302 (2010).
83. Krasileva KV *et al*, *PLoS ONE* 6:e28765 (2011).
84. Goritschnig S *et al*, *PLoS Genet* 8:e1002502 (2012).
85. Kuang H *et al*, *Plant Cell* 16:2870 (2004).
86. Schneeberger K *et al*, *Genome Biol* 10:R98 (2009).
87. Turner EH *et al*, *Nat Methods* 6:315 (2009).
88. Galvão VC *et al*, *Plant J* 71:517 (2012).
89. Li Y *et al*, *PNAS* 107:21199 (2010).
90. Baxter I *et al*, *PLoS Genet* 6:e1001193 (2010).
91. Koornneef M *et al*, *Annu Rev Plant Physiol Plant Mol Biol* 55:141 (2004).
92. Kliebenstein DJ *et al*, *Genetics* 172:1179 (2006).
93. Keurentjes JJ *et al*, *Nat Genet* 38:842 (2006).
94. Keurentjes JJ *et al*, *PNAS* 104:1708 (2007).
95. Becker C *et al*, *Nature* 480:245 (2011).
96. Listgarten J *et al*, *Nat Methods* 9:525 (2012).
97. Lippert C *et al*, *Nat Methods* 8:833 (2011).
98. Kam-Thong T *et al*, *Human heredity* 73:220 (2012).
99. Kam-Thong T *et al*, *Eur J Hum Genet* 19:465 (2011).
100. Kam-Thong T *et al*, *Bioinformatics* 27:i214 (2011).
101. Consonni C *et al*, *Nat Genet* 38:716 (2006).
102. Monaghan J, Zipfel C, *Curr Opin Plant Biol* 15:349 (2012).
103. Bentsink L *et al*, *PNAS* 103:17042 (2006).
104. Charlesworth D, Willis JH, *Nat Rev Genet* 10:783 (2009).
105. Kusaba M *et al*, *Plant Cell* 13:627 (2001).
106. Ross-Ibarra J *et al*, *PLoS ONE* 3:e2411 (2008).
107. Pyhäjärvi T *et al*, *Am J Bot* 99:1314 (2012).
108. Fobis-Loisy I *et al*, *Plant Cell Rep* 26:745 (2007).
109. Xie W *et al*, *PNAS* 107:10578 (2010).
110. Andolfatto P *et al*, *Genome Res* 21:610 (2011).
111. Plotnikova JM *et al*, *Mycologia*:1009 (1998).
112. Slusarenko AJ, Schlaich NL, *Mol Plant Pathol* 4:159 (2003).
113. Thines M *et al*, *Persoonia* 22:123 (2009).
114. Baxter L *et al*, *Science* 330:1549 (2010).
115. Spanu PD *et al*, *Science* 330:1543 (2010).
116. Kemen E *et al*, *PLoS Biol* 9:e1001094 (2011).
117. Coates ME, Beynon JL, *Annu Rev Phytopathol* 48:329 (2010).
118. Allen RL *et al*, *Science* 306:1957 (2004).
119. Holub EB *et al*, *MPMI* 8:916 (1995).
120. Holub EB, Beynon JL, *Adv Bot Res* 24:227 (1997).
121. Gollner K *et al*, *New Phytol* 177:725 (2008).
122. Jakob K *et al*, *MPMI* 15:1195 (2002).
123. Araki H *et al*, *PNAS* 103:5887 (2006).
124. Goss EM, Bergelson J, *Evolution* 60:1562 (2006).
125. Goss EM, Bergelson J, *Oecologia* 152:71 (2007).
126. Goss EM *et al*, *Genetics* 169:21 (2005).
127. Knief C *et al*, *Microb Ecol* 60:440 (2010).
128. Knief C *et al*, *ISME J* 4:719 (2010).
129. Bulgarelli D *et al*, *Nature* 488:91 (2012).
130. Lundberg DS *et al*, *Nature* 488:86 (2012).
131. Bomblies K *et al*, *PLoS Genet* 6:e1000890 (2010).
132. Wang Y, Qian PY, *PLoS One* 4:e7401 (2009).
133. Cole JR *et al*, *Nucleic Acids Res* 37:D141 (2009).
134. McDonald D *et al*, *ISME J* 6:610 (2012).
135. Huson DH, Mitra S, *Methods Mol Biol* 856:415 (2012).
136. Melnikov A *et al*, *Genome Biol* 12:R73 (2011).
137. Namiki T *et al*, *Nucleic Acids Res* 40:e155 (2012).
138. Peng Y *et al*, *Bioinformatics* 27:i94 (2011).
139. Li R *et al*, *Genome Res* 20:265 (2010).
140. Lozupone C *et al*, *ISME J* 5:169 (2011).
141. Arumugam M *et al*, *Nature* 473:174 (2011).
142. Qin J *et al*, *Nature* 490:55 (2012).