Invited Talk

## Genomic Variation and Incipient Speciation in Arabidopsis thaliana

Detlef Weigel

Max Planck Institute for Developmental Biology, Germany
and Salk Institute for Biological Studies, La Jolla, USA.

Comprehensive polymorphism data are a prerequisite for the systematic identification of sequence variants affecting phenotypes. In the first part of my talk, I will discuss our efforts to provide a whole genome resource for the study of population level evolutionary processes in an experimentally tractable, multicellular organism, Arabidopsis thaliana. To this end, we have collaborated with Kelly Frazer and colleagues at Perlegen Sciences, and hybridized genomic DNA of 20 strains to custom microarrays that tile all possible single nucleotide polymorphisms (SNPs) along the entire genome with close to one billion (109) different oligonucleotides. The analysis of SNP distribution and haplotype maps is being carried out in collaboration with the groups of Bernhard Schölkopf (MPI for Biological Cybernetics), Gunnar Rätsch (Friedrich Miescher Laboratory), Daniel Huson (University Tübingen), Joe Ecker (Salk Institute), and Magnus Nordborg (USC). Using novel analysis methods, we identified up to 1.1 million non-redundant SNPs at various levels of precision. In addition, we predicted nearly 5% of the genome to be highly polymorphic or deleted in at least one strain. These data allow for the first time a systematic description of the types of genes that harbor major changes (e.g., stop codons or whole gene deletions) in wild populations. Although major changes are frequent, allele frequency patterns indicate that they are often associated with a fitness cost. Disease resistance (R) genes are found to be the most polymorphic class of genes.

Through our work on natural variation, we have also become involved in more general questions of species-wide evolution. It has long been suggested that post-zygotic hybrid incompatibility between closely related species arises as a by-product of deleterious interactions between genes that have diverged since the most recent common ancestor. In animals, several such gene pairs have been identified in interspecies crosses, but it is not yet known whether they play only a role in maintaining species boundaries, or whether they are also important in establishing barriers to gene flow. To understand the mechanisms underlying nascent incompatibilities, we performed an extensive survey for hybrid incompatibilities within A. thaliana. We identified numerous independent F1 incompatibilities with a range of phenotypically related abnormalities. Each case is attributable to two to three epistatic loci. A common autoimmune mechanism--activation of pathogen responses in the absence of pathogens--underlies the majority of incompatibilities. Moreover, in a collaboration with Jeff Dangl (UNC), we have found that higher disease resistance correlates with incompatibility phenotypes, suggesting a fitness trade-off. Detailed characterization of one hybrid interaction identified a disease resistance (R) gene variant as causal for the incompatibility phenotype. R genes constitute the fastest evolving gene family in plants, suggesting that such incompatibilities arise frequently as a by-product of natural selection.