Experimental Studies of Complex Molecular Systems

Laura F. Landweber

I study the origin of novel genetic systems in single-celled eukaryotes, focusing on the interplay between molecular evolution and genetic computation. We have shown that their surprisingly sophisticated variations on DNA and RNA processing create an imaginative playground for genome architecture and genetic systems. Some of their pathways erode the notions of a gene (e.g. scrambled genes and RNA editing) and even Mendelian inheritance, reminding us that a genome sequence can be a far cry from knowledge of its products. Genome-wide DNA rearrangements occur in many diverse eukaryotes, but their extreme exaggeration in ciliates makes them an ideal model system to study epigenetic phenomena. Our laboratory is currently focused on the mechanism and evolution of this remarkable process. By combining molecular, evolutionary, theoretical, and synthetic biology, we recently discovered an RNA-guided epigenetic mechanism underlying programmed genome rearrangements (Nowacki *et al.* 2008 *Nature*), as well as the transitional, or putative intermediate, forms that this complex system passes through during both development (Möllenbeck *et al. PLoS ONE,* 2008) and evolution (Chang *et al.* 2005 *PNAS*) and its intrinsic capacity for solving computational problems. Nowacki *et al*. 2008 also unveiled a new role for RNA, a new class of noncoding RNAs, and a new mechanism for epigenetic (re)wiring of cellular programs, which may be general.

Our previous work has also made fundamental contributions to other areas of research, including creating the first "RNA computer" (Faulhammer *et al.* 1999 *PNAS*), which held the record for molecular computation for several years, visualizing natural biological systems as computation (e.g. Landweber and Winfree 2002), and turning the study of the origin of the genetic code into a quantitative, rigorous testing ground (Knight and Landweber 1998, 2000 and several other publications with Knight or Freeland), disproving that the genetic code was ever a frozen accident. Because ciliates bear the richest diversity of alternative genetic codes, they even helped us illuminate the rules underlying transformation of the genetic code itself (Lozupone *et al.* 2001 *Curr Biol.*; Liang *et al*. 2005 *JME*). More recently, several studies in my lab examined other aspects of protein translation using comparative genomics (Liang et al. 2005 *Genome Biol*.; Liang and Landweber 2006 *Genome Res.*) and structural data (Liang and Landweber and Liang *et al.*, both 2005 in *RNA*) to examine its early evolution from the RNA world.

The Mechanism of Gene Unscrambling

A recent development in my lab has been a close examination of the molecular basis for DNA rearrangement in ciliates. During development of the somatic macronucleus, *Oxytricha trifallax* destroys an unprecedented 95% of its germ-line, severely fragmenting its chromosomes, and then unscrambles hundreds of thousands of remaining fragments by permutation or inversion. Programmed DNA deletion, inversion, and permutation events provide the main experimental system we are using to explore the mechanism underlying complex genome rearrangements. Massive rearrangement is required to construct a set of gene-sized “nanochromosomes” for expression in *Oxytricha*'s somatic

macronucleus. These DNA acrobatics include programmed loss of intergenic DNA as well as internal eliminated sequences (IESs) within genes, accompanied by assembly of protein-coding segments (macronuclear destined sequences, or MDSs) in their correct order from germline segments sometimes present in a scrambled, permuted order. Pointer sequences (word matches) at the junctions between coding and noncoding sequences allow reconstruction of intact genes during development.

Based on our current work, the molecular mechanisms of gene unscrambling are beginning to come to light. We recently demonstrated an epigenetic mechanism for DNA unscrambling in *Oxytricha* (Nowacki *et al.* 2008 *Nature*). Our model proposes the existence of RNA templates, complete cached copies of the previous generation's reorganized genome, and that these molecules guide assembly of cryptic germline fragments in a correct order and orientation (Angeleska *et al.* 2007 *J. Theor. Biol.*). Nowacki *et al.* 2008 offers several lines of strong empirical evidence for the transient existence and action of these long, double-stranded RNA templates. We first successfully disrupted gene unscrambling in *Oxytricha* by either microinjecting or feeding the cells bacteria expressing specific double-stranded RNA molecules. This leads to aberrant gene unscrambling in the progeny, blocking the process completely or producing new, incorrect DNA patterns in the targeted gene. This new and exciting line of experiments has provided us with the first experimental opportunity to parse the specific steps in this complex system and has permitted us to develop *Oxytricha* as a powerful model system. Together with Vikram Vijayan '07, whose senior thesis received highest distinction in Electrical Engineering, we microinjected synthetic DNA or RNA templates in the developing cell nuclei and showed that these artificial molecules can orchestrate new genome rearrangements in *Oxytricha*, reprogramming the DNA rearrangement pathway for multiple generations. This strongly suggests that single- or double-stranded RNA molecules guide genome rearrangement (Nowacki, Vijayan, et al. *Nature,* 2008), supporting our epigenetic model for the molecular basis of information transfer in genome rearrangement. More broadly, it reveals a new role for RNA in biology, a new class of noncoding RNAs (template RNA), and a new mechanism for epigenetic wiring and rewiring of cellular programs, along with a pathway for transmission of an acquired genetic state. We are also beginning to probe the biochemical mechanism of DNA rearrangement, which we think is dependent on a group of germline-limited transposase genes, based on widespread disruption of genome rearrangement after simultaneous RNAi knockdown of three candidate transposase families (Nowacki *et al., Science*, in revision).

The Origin and Evolution of Scrambled Genes

We have proposed two different step-wise models for how scrambled genes have become extensively fragmented and permuted over evolutionary time (Chang *et al.* 2005 *PNAS*; Wong and Landweber 2006 *Mol Biol Evol*). In a related study we found that coding and non-coding sequences can even exchange during recent gene evolution (Möllenbeck *et al.* 2006 *J. Mol Evol*). We have also discovered and characterized new scrambled genes in detail (Kuo *et al.* 2006 *Mol Biol Evol*; Chang *et al.* 2006 *Gene*). The permuted coding segments for two scrambled genes are surprisingly intertwined in the germline of one organism. But the record for the most scrambled germline locus to date is a surprisingly

unconventional set of ribosomal RNA genes in *Oxytricha,* fragmented into approximately *150 scrambled pieces* present on 4 or more unlinked loci in the germline (Maquilan *et al*., in preparation). In an essay and subsequent correspondence in *Science* (Landweber 2007 and 2008) I discuss various rationales for the evolutionary origin of "genomes in pieces" and whether the presence of such a scattered genetic architecture might actually confer a selective benefit to some organisms.

***Oxytricha* Genomics**

Because thousands of genes, or approximately 20-30% of the germline genome of spirotrichous ciliates, are estimated to be scrambled, we are conducting a genome sequencing project for *Oxytricha trifallax*. In collaboration with the NHGRI-funded Genome Center at Washington University, we are sequencing the somatic (macronuclear) genome of *Oxytricha trifallax* to high coverage and the germline (micronuclear) genome to 2X coverage (genome sizes approximately 50 Mb and 1 Gb, respectively). The macronuclear draft sequence is approaching conclusion. High quality sequencing of *Oxytricha*'s macronuclear genome elegantly avoids the problem of sequencing the 95% of the micronuclear genome that is non-coding DNA and provides the sequence in the form of gene-sized units with associated regulatory regions of only ~100 bp. We have created tools along the way that benefit the entire bioinformatics and comparative genomics community, such as "BLASTO", a webserver for searching sequence similarity among orthologous groups (Zhou and Landweber, 2007) which has over 4500 users so far. These genomic resources will pave the way for future studies of the more complex genetic architecture of the micronuclear genome. Because our estimates suggest that as much as one-third of the germline genome (containing several thousand genes) is scrambled, on the order of approximately 10,000 scrambled genes, this presents us with a computation challenge for the next few years. Such a high resolution genomic analysis will lead to a broader understanding of the complex and dramatic genome rearrangements that can accumulate in an organism over evolutionary time.