

Katy L. Simonsen

Research Statement

October 3, 2005

I. CREATIVE ENDEAVOUR, RESEARCH, SCHOLARSHIP

Dr. Simonsen's research is at the interface of statistics, genetics, applied probability, mathematical modeling, and scientific computation. The major objective of Dr. Simonsen's multidisciplinary research is to understand and include the forces which contribute to genetic variability in natural populations, and the consequences of that variability on future natural populations and our ability to interpret data from present-day organisms. Natural populations of organisms such as humans are produced by naturally occurring matings without experimental intervention. Understanding the way DNA sequences vary among organisms is essential to our ability to detect and locate genes for specific traits and diseases. The impact of this can be felt in highly prevalent complex diseases such as diabetes and mental illness, and in disease management strategies such as pharmacogenomics. Phenomena such as natural selection, genetic recombination, population structure, mutation, and random genetic drift are examples of major forces that interact to create, maintain, and eliminate genetic variability. Dr. Simonsen's research is centered on developing novel mathematical and computational models of such phenomena, and uses these models to construct reliable procedures for estimating parameters and testing hypotheses about genes, traits, organisms, and their natural history.

A. Discussion of Research

Specific projects include the following¹.

1. Testing the neutral model of molecular evolution

The neutral theory of molecular evolution proposes that most DNA sequence variation within populations is selectively neutral and the result of chance variation, while a relatively small fraction of genomes are subject to natural selection or other directive forces. Many hypothesis tests have been proposed for detecting deviations from the neutral model of molecular evolution that underlies this theory[7]. This is of importance and highly relevant to understanding how disease and trait genes are passed from generation to generation. Specifically, population geneticists wish to test specific regions of a genome to see whether neutrality holds. To address this issue Dr. Simonsen used computer simulation to compare the power of different testing procedures that detect complex alternatives such as natural selection, population structure, and population size fluctuations [10]. This comparison revealed the relative merits of each testing procedure and exposed important flaws common to all existing procedures.

According to the neutral model, mutations in DNA sequences occur at random, but at a constant rate over time. The null hypothesis of neutrality is further complicated by the fact that the mutation rate is unknown, and thus this rate acts as a nuisance parameter. Computer simulations of the null hypothesis allow for the construction of critical regions for hypothesis tests, but these cannot be performed without specifying this nuisance parameter. At the time (1993), existing literature tended to use a single point estimate for this mutation rate parameter, without taking into account the underlying uncertainty. Dr. Simonsen was the first to point out that since this point estimate was not based on a sufficient statistic for the parameter, statistical theory dictates that information about the parameter was lost in the process. The result was misspecification of the critical regions,

¹Citations here refer to the publications list B.

which in turn affected both the power and false positive rate of the hypothesis tests. Dr. Simonsen tackled this limitation by devising a novel way to construct confidence intervals for the mutation rate nuisance parameter, and by adapting a method from the statistics literature to construct new, accurate, critical regions for existing test statistics. She accomplished this difficult task using computer simulation and by maximizing over a confidence interval for the mutation rate nuisance parameter.

Capitalizing on the novel approach for constructing more accurate critical regions for testing, Dr. Simonsen adapted and developed coalescent methods to simulate DNA sequences under neutrality and various alternative hypotheses such as strong selection, balancing selection, and population size change. Coalescent methods are based on a stochastic process going backwards in time and allow the simulation of the ancestral tree representing the relationship among sampled individuals (e.g., siblings). This type of simulation is much more powerful and efficient than evolving a large population forwards in time, and then taking a small sample. Using these simulation methods to compare the power of several test statistics in these alternative situations, Dr. Simonsen exposed many interesting results. Specifically, the most commonly used test in the literature, Tajima's D , was found to be not very powerful against most alternatives. Thus, Tajima's D and the related tests are not able to detect anything but the most extreme departures from neutrality.

The main impact of Dr. Simonsen's research and results instigated several important changes to the field. First, the aforementioned tests are now interpreted in the literature within the framework of their statistical power. Prior to publication of this work, it was common to see a non-rejection with Tajima's D presented as evidence in favor of neutrality; afterwards, it became more common (correctly) to interpret such results as a lack of evidence for or against neutrality. Second, more powerful statistical tests were developed and continue to be developed as it was realized the existing tests were not sufficiently specific. Third, it has become standard practice, since Dr. Simonsen's exposure of the problem, to perform power studies when a new test is proposed, and to use simulation to determine critical regions for tests. Dr. Simonsen's publication in 1995 has led to a revolutionary change that has exposed limitations of previous methods as applied to the neutral model of evolution and has allowed the field to move substantially forward.

2. Coalescence with recombination

As mentioned previously, coalescent models are a type of stochastic process that are employed to model the distribution of the ancestry of a random sample of DNA sequences. Understanding the behaviour of random samples of DNA sequences allows understanding of evolutionary events, and empowers future scientific discovery by leveraging information. Most theoretical and simulation work involving coalescent models has concentrated on the simple no-recombination model in which all loci are completely linked and have exactly the same ancestry. In other words, the level of realistic variation that occurs in the formation of a gamete is limited. The only existing model of coalescence that included recombination allowed only a single chromosome-wide recombination rate to be specified, so that each recombination event occurred at a new, random, location. Biologically, this assumption is unrealistic because recombination rates vary widely along chromosomes, and because genetic marker data is typically collected at fixed chromosomal positions. Dr. Simonsen developed a novel coalescent model that included recombination by using a more realistic scenario of a fixed number of loci separated by specific recombination rates. This in turn enabled multiple recombination events to occur at the same position [9].

Dr. Simonsen's approach based on a Markov Chain model traces the possible ancestries of a set of sequences by traveling backwards in time through its history in a probabilistic manner. Starting in a specific "state", the model proceeds through a series of recombination and coalescent "events", which allow the model to move to different states and whose relative probabilities are determined by the present state. This model is especially applicable to the important case of two sequences and two loci (e.g., two disease genes or a disease gene and a genetic marker). Dr. Simonsen mathematically derived a number of properties for this model and proved its convergence. One important contribution was to provide the exact probability distribution of the number of recombination events in the history of a sample, a result that had previously only been approximated.

As exposed by the specific case of two sequences and two loci, when the number of sequences or loci exceeds two, the number of states in the Markov Chain increases so quickly that a theoretical analysis and result is not mathematically tractable. Dr. Simonsen addressed this issue by using the Markov Chain as the foundation for constructing a computer algorithm to sample from the set of possible ancestries. This in turn allowed her to simulate data sets of DNA sequences with specific recombination rates and correspondingly realistic levels of genetic variation and linkage disequilibrium.

Because mutations are passed on through ancestral lineages, modeling ancestry allows the progression of genetic variation over time to be modeled. Furthermore, because of genetic recombination, different sites of a DNA sequence can have different ancestries. Failure to acknowledge this genetic variation produces results that are biologically irrelevant. Dr. Simonsen's research in this area changed the understanding of how genes and DNA sequences are passed from generation to generation by permitting models that incorporate varying levels of recombination across any genome.

3. Efficient simulation algorithm for coalescence with recombination

In an effort to continually improve and focus her research, Dr. Simonsen developed, with Computer Science M.Sc. student Daniel Noland, a much more efficient algorithm to simulate the previously mentioned Markov Chain [13] than was originally employed for her initial work [9]. The new algorithm requires computational time and memory which is only quadratic in the number of loci (m^2) and linear in the number of sequences, whereas the original algorithm was exponential (2^m). This is a significant computational improvement over the initial implementation and is a joint collaboration with the Research Computing Division of ITaP. The improved algorithm permits genome-scale simulations, namely where loci number in the thousands. This algorithm was named Argos for Ancestral Recombination Graph Optimized Simulator.

The Argos algorithm was implemented into publicly available software in conjunction with the Research Computing Division of ITaP. It was integrated into a suite of programs known as SIMCORE for SIMulating COalescence with REcombination and made publicly available.

4. Association between traits and genetic markers

Locating genes in any genome of interest lends information to the inner workings of both organisms and diseases. The biological restrictions (e.g., recombination) that are involved in the statistical approaches for locating genes often limits success. Dr. Simonsen analyzed the effect of recombination rate and marker frequency on statistical methods for gene location, using a computer-based method described in [9]. This work examined the Transmission-Disequilibrium Test (TDT), used

to detect and locate genes associated with human genetic disease [6]. Combining her previous work in coalescent theory with locating genes, Dr. Simonsen performed a computer-based coalescent simulation to generate large numbers of datasets with realistic levels of genetic variation/linkage disequilibrium. Various multiple testing procedures were assessed to compare their power to detect genes associated with disease traits.

The problem of locating genes associated with traits and diseases is complicated by the necessity of controlling false positive rates or false discovery rates in multiple testing. Dr. Simonsen conducted a simulation study of a backcross experimental design to explore combinations of test statistics and multiplicity corrections [5, 24, 25]. Three new tests based on order statistics were proposed: Order, ELM (Elimination of Linked Markers), and ERP (Eliminate and Re-Permute). These were compared with three existing tests: simple locus-wise, maximum order statistic, and CET (Conditional Elimination Test). Two new multiplicity corrections (geometric and spend-as-you-go) were proposed based on the idea of an alpha-spending function. A comparison of these two novel approaches to four common approaches: Bonferroni, Benjamini-Hochberg, Benjamini-Yekutieli, and no correction was performed. The six tests were examined for their ability to detect multiple trait loci when combined with the six different multiplicity corrections. The new approaches tended to outperform existing ones especially in their ability to identify all trait loci (as opposed to finding just one). This investigation revealed certain test and correction combinations that either severely over-controlled or failed to control false discovery rates at the desired level. The impact of this lends itself to a better understanding of the performance of current methods. The result of this research included the development of software for performing data analysis using all combinations of approaches. This software is publicly available at <http://www.stat.purdue.edu/~simonsen/AlphaWisely/>. We used this approach and software to reanalyze a published dataset and identified a number of quantitative trait loci for oat vernalization not found in the original analysis.

Many statistical approaches for locating genes rely on likelihood functions. In order to use likelihood-based statistical inference methods, the relationship between markers, genes, and trait values must be described mathematically via a model. For a single marker-gene pair, the relationship is simple to describe under a known mating structure, but the problem rapidly becomes notationally and mathematically complex with more linked loci. Dr. Simonsen developed a very general probability model and matrix notation to describe these inheritance patterns for arbitrary numbers of loci in experimental populations [14]. The implementation of a matrix formulation allows the entire genome to be described by the relationship between all markers and genes for any experimental design. The framework is general enough to incorporate any type of genetic model, without requiring simplifying assumptions such as additivity or dominance. Furthermore, the trait model is completely general and can include binary or quantitative traits. The impact of this general result allows one to locate genes relative to any number of markers in any experimental system. Dr. Simonsen automated the calculation of the probability model for any given number of marker-gene pairs, which in turn permits a straightforward and automated likelihood calculation to be used to estimate parameters and test hypotheses [2, 27]. The continuing long-term goals of this research include using this model formulation to expand and improve gene mapping procedures for both binary and quantitative traits, as well as extending this formulation to include natural populations.

5. Traits in Natural Populations

Combining the coalescent model approach with methods to locate trait loci in natural populations unifies two of Dr. Simonsen's major research areas. Funded by the National Science Foundation, this work with postdoctoral associate Dr. Koen Verhoeven and co-Principal Investigator Dr. Lauren McIntyre uses a novel tree-based search method to detect loci associated with traits in natural populations.

Identifying such loci in natural populations is difficult because the frequencies of markers and traits are the result of random sampling instead of a designed experiment. The sampling distribution of these data and associated likelihood formulae are essentially impossible to pinpoint analytically. Fortunately, the previously mentioned efficient algorithm for coalescence with recombination developed by Dr. Simonsen is ideally suited to simulating marker data that has the same properties as real natural population data. Large scale simulations of such data are being used to evaluate procedures for identifying marker-trait associations.

As part of this research, Dr. Simonsen has developed a modular suite of programs called SIMCORE for the purpose of providing a complete understanding of data simulation, model selection, and power analysis. SIMCORE allows for the efficient simulation of coalescent trees under the assumption of recombination using the Argos algorithm, and allows one to optionally adjust the evolutionary trees to accommodate changes in population size. It then places mutations on trees according to one of several possible mutation models to produce sequence data. From this sequence data, haplotype blocks are identified and linkage disequilibrium calculated, along with marker frequencies and other summary statistics. Associated trait data are generated according to a specified genetic model based on the genotypes found in the sequence data. The networks comprising sufficient statistics for the sequence data are generated and used to identify recombination breakpoints based on phylogenetic incompatibility. Finally, a tree-based model selection procedure is performed to identify the underlying genetic model (i.e. QTL) and the results are summarized in terms of power, false discovery rates, etc. The software has been made publicly available (<http://www.stat.purdue.edu/~simonsen/SIMCORE/>), and SAS Institute has plans to incorporate much of this software as a stand-alone simulation procedure "PROC SIMCORE" in SAS/Genetics.

6. Understanding Haplotype Blocks

Recent genomic studies have detected a block-like structure of linkage disequilibrium in the human genome, which is thought to be caused by spatial clustering of recombination events. An enormous effort is underway to exploit this haplotype block structure for gene mapping studies (the International HapMap project). Dr. Simonsen (in collaboration with Dr. Verhoeven) employed coalescent simulations to demonstrate that the relation between historic recombination events and haplotype blocks is much weaker than is commonly assumed [3]. Moreover, the simulations give insight into why the relationship is weak. These results provide fundamental insight into the relationship between historic recombination and linkage disequilibrium patterns. They stress the power of stochastic events (as opposed to clustered recombinations) to shape haplotype blocks, which limits the envisioned haplotype tagging approach to gene mapping. The approach is based on the assumption that the structure is consistent among and within different populations, which is likely not the case if the blocks develop stochastically.

Although sequence data is based on an underlying set of coalescent trees, it is usually not possible to completely reconstruct those trees based only on sequence data, since branches without

mutations cannot reliably be inferred. However, there is a tree-like or network structure that can be constructed from phylogenetically compatible sequences, and this network is the sufficient statistic for the data. Dr. Simonsen developed an algorithm to divide sequences into phylogenetically compatible blocks, and then to construct the appropriate network for each block. The foundation of statistical theory dictates that the best testing and estimation procedures are based on sufficient statistics; thus this set of networks is the most efficient starting point for any statistical analysis of sequence data.

7. Construction and Transportation Engineering

Although completely outside Dr. Simonsen's research focus of statistical genetics, she has engaged in a number of consulting projects related to statistical analysis of safety in Civil Engineering. As a secondary interest it has allowed Dr. Simonsen to refine her statistical consulting skills, and ties in well with her teaching interests in Applied Statistics.

With Ph.D. students Vandana Patidar, Alberto Figueroa-Medina, Samantha Islam, and Peter Savolainen (Transportation Engineering) Dr. Simonsen is involved in research problems that are related to the statistical analysis of data for the identification of factors associated with traffic accidents, accident severity, and fatality rates. The results of this work have the potential for great impact on society by reducing accidents and fatalities.

With graduate student Javier Irizarry (Construction Engineering) and his advisor Dr. Dulcy Abraham, Dr. Simonsen was instrumental in analyzing observational data as related to safety at construction sites [11]. These results showed the use of safety measures to be associated with only a negligible reduction in productivity. Based on the success of this work Dr. Simonsen and her collaborators have two outstanding proposals with the National Institute of Occupational Safety and Health. Among the objectives of this work are to conduct randomized controlled trials to help elucidate the causal relationship between safety and productivity in steel construction. This work has the potential for great impact in steel construction by reducing worker injuries while maintaining productivity.

B. Publications

To give an idea of the ratings of journals, the Journal Impact Factor (JIF) (www.isiknowledge.com) has been included for each journal. It must be recognized that the range of Impact Factors varies widely by field due to the number of journals, the size of the field, and their citation rates. For example, the highest JIF in Civil Engineering is 1.481 (*J. Hydrology*), in Statistics & Probability is 5.724 (*Bioinformatics*), in Ecology is 12.938 (*Trends in Ecology and Evolution*), and in Genetics & Heredity is 24.695 (*Nature Genetics*). In addition, a journal may be listed and ranked in several categories. Therefore, a comparison of JIF across disciplines is both problematic and challenging for any interdisciplinary researcher. To offset the cultural insensitivity of JIF, the number of times an article has been cited has also been provided when available. It can be seen from the following table that while Dr. Simonsen publishes in a variety of subject areas, the journals in which her work appears are at the top of their respective fields.

Journal Name	Impact Factor	Rank	Tier	Peer Group Category
Trends in Ecology and Evolution	12.938	1/107	Top	Ecology
		1/33	Top	Evolutionary Biology
		5/120	Top	Genetics & Heredity
Molecular Biology and Evolution	6.355	14/120	Top	Genetics & Heredity
		4/33	Top	Evolutionary Biology
Genetics	4.138	25/120	Top	Genetics & Heredity
Genetic Epidemiology	3.038	12/93	Top	Public, Environmental & Occupational Health
Oikos, A Journal of Ecology	2.901	19/107	Top	Ecology
Theoretical Population Biology	2.481	26/107	Top	Ecology
Journal of Construction Engineering and Management	0.284	30/79	Middle	Engineering, Civil

1. Refereed

- [1] B. Munneke, K. A. Schlauch, K. L. Simonsen, W. D. Beavis, and R. W. Doerge, “Adding confidence to gene expression clustering,” *Genetics*, vol. 170, pp. 2003–2011, August 2005.
- [2] C. J. Coffman, R. W. Doerge, K. L. Simonsen, K. M. Nichols, C. K. Duarte, R. D. Wolfinger, and L. M. McIntyre. “Model Selection in Binary Trait Locus Mapping”. *Genetics*, vol. 170, pp. 1281–1297, July 2005.
- [3] K. J. F. Verhoeven and K. L. Simonsen. “Genomic haplotype blocks may not accurately reflect spatial variation in historic recombination intensity”. *Molecular Biology and Evolution*, vol. 22, pp. 735–740, 2005. (Times cited: 1)
- [4] K. J. F. Verhoeven, K. L. Simonsen, and L. M. McIntyre. “Implementing false discovery rate control: increasing your power”. *Oikos, A Journal of Ecology*, vol. 108, no. 3, pp. 643–647, 2005. (Times cited: 2)
- [5] K. L. Simonsen and L. M. McIntyre, “Using alpha wisely: Improving power to detect multiple QTL,” *Statistical Applications in Genetics and Molecular Biology*, vol. 3, no. 1, article 1, 2004. <http://www.bepress.com/sagmb/vol3/iss1/1> (Times downloaded: 224)
- [6] L. M. McIntyre, E. R. Martin, N. L. Kaplan, and K. L. Simonsen, “Circumventing multiple testing: A multilocus Monte Carlo approach to testing for association,” *Genetic Epidemiology*, vol. 19, no. 1, pp. 18–29, 2000. (Times cited: 32)
- [7] M. L. Wayne and K. L. Simonsen, “Statistical tests of neutrality in the age of weak selection,” *Trends in Ecology and Evolution*, vol. 13, pp. 213–253, 1998. (Times cited: 36)

- [8] K. L. Simonsen, N. L. Kaplan, and E. R. Martin, “A Monte-Carlo permutation approach to choosing an affection status model for bipolar affective disorder,” *Genetic Epidemiology*, vol. 14, pp. 681–686, 1997.
- [9] K. L. Simonsen and G. A. Churchill, “A Markov Chain model of coalescence with recombination,” *Theoretical Population Biology*, vol. 52, no. 1, pp. 43–59, 1997. (Times cited: 7)
- [10] K. L. Simonsen, G. A. Churchill, and C. F. Aquadro, “Properties of statistical tests of neutrality for DNA polymorphism data,” *Genetics*, vol. 141, pp. 413–429, 1995. (Times cited: 197)
(citations as of 9/05)

2. In Press

- [11] J. Irizarry, K. L. Simonsen, and D. M. Abraham. “Effect of Safety and Environmental Variables on Task Durations in Steel Erection”. Accepted to *Journal of Construction Engineering and Management*, 2005.
- [12] J. Irizarry, K. L. Simonsen, and D. M. Abraham. “Safety and environmental variables: effect on steel erection task durations”. Accepted to *3rd International Structural Engineering and Construction Conference* Shunan, Japan, September 2005 (refereed conference paper).

3. Submitted

- [13] K. L. Simonsen and D. A. Noland, “Argos: An efficient algorithm for simulating the coalescent with recombination”. Submitted to *Bioinformatics*, 2005.

4. In Preparation

- [14] K. L. Simonsen. “A general probability model for the inheritance of binary traits”. To be submitted to *Theoretical Population Biology*, 2005.

5. Software

- [15] SIMCORE: Software for SIMulating COalescence with REcombination, sequence generation, trait generation, and data analysis using the Argos algorithm. (<http://www.stat.purdue.edu/~simonsen/simcore/>). Also in the process of being adopted by SAS Institute for incorporation into SAS/Genetics software, as PROC SIMCORE.
- [16] AlphaWisely. Software for analyzing genetic data to detect QTL using multiple FDR-controlling procedures (<http://www.stat.purdue.edu/~simonsen/AlphaWisely/>).

6. Published Research Abstracts

- [17] K. L. Simonsen, D. A. Noland and C. Le. “An efficient algorithm for simulating coalescence with recombination”. *Computing Science and Statistics*, vol. 36, 2004.
- [18] K. L. Simonsen, “Genome Data: Finding Trait Genes Using a Dense Marker Map”, *Computing Science and Statistics*, vol. 31, p. 241, 1999.
- [19] M. L. Wayne, L. G. Harshman, K. L. Simonsen, and A. Salmon, “Statistical issues in functional genomics: temperature-dependent gene expression in *Drosophila melanogaster*”. *Nature Genetics: The microarray meeting*, vol. 23 supplement, p. 81, 1999.

7. Non-Refereed Research Publications

- [20] G. Casella, G. A. Churchill, and K. L. Simonsen, “Sampling based methods for the estimation of DNA sequence accuracy,” Tech. Rep. BU-1138-M, Cornell University, Statistics Department, 1995.
- [21] K. L. Simonsen, 2003. “A general probability model for the inheritance of binary traits”. Technical Report #TR-03-04, Purdue University Statistics Department, 2003.

8. Posters

- [22] C. J. Coffman, R. W. Doerge, K. L. Simonsen, K. Nichols, C. Duarte, R. Wolfinger and L. M. McIntyre. “Model Selection in Binary Trait Locus Mapping”. Gordon Research Conference on Quantitative Genetics and Genomics, Ventura, CA, February 2005.
- [23] Cynthia J. Coffman, Krista Nichols, Christine Woods, Wendy Czika, Katy L. Simonsen, Russell D. Wolfinger, Lauren M. McIntyre “Model Selection Strategy for Multiple Binary Trait Loci” Joint Statistics Meeting, San Francisco, CA, 2003.
- [24] L. M. McIntyre, Y. Qin and K. L. Simonsen. “Multiple Testing: An Ordered Approach” Plant and Animal Genome X, San Diego CA, 2002.
- [25] L. M. McIntyre, K. L. Simonsen, Y. Qin “Multiple Testing and Genetic Marker Data: An Ordered Approach.” Joint Statistics Meeting, New York NY, 2002.
- [26] K. L. Simonsen, and L. M. McIntyre. “Multiple Testing Issues for Linked Genetic Markers”. Eastern North American Regional Meeting of the International Biometrics Society. Charlotte, NC, March 2001.
- [27] C. J. Coffman, R. W. Doerge, K. L. Simonsen, and L. M. McIntyre. “Detection and Localization of Multiple Binary Trait Loci in Experimental Populations.” Eastern North American Regional Meeting of the International Biometrics Society; Charlotte, NC, March 2001.
- [28] L. M. McIntyre, E. R. Martin, K. L. Simonsen and N. L. Kaplan. “Circumventing the Multiple Testing Problem in Searching for Disease Genes”. Sixth International Purdue Symposium on Statistics; Purdue University, West Lafayette IN, June 1998.

- [29] L. M. McIntyre, E. R. Martin, K. L. Simonsen and N. L. Kaplan. “The TDT for Multiple Loci” Meeting of the American Society for Human Genetics; San Francisco, CA, October 1997.
- [30] K. L. Simonsen. “How Powerful is Tajima’s D?” New England Molecular Evolution Meeting; Harvard University, Cambridge, MA, November 1994.

C. Invited Lectures

1. *A Polynomial Algorithm for Simulating Coalescence with Recombination*, Probability and Statistics Seminar, Indiana University, Bloomington (April 5, 2004).
2. *Life at the Interface: My Multidisciplinary Career in Science*, Indiana University, Bloomington (April 5, 2004).
3. *An Efficient Algorithm for Simulating Coalescence with Recombination*, Bioinformatics Seminar, Purdue University (Feb. 17, 2004).
4. *A Polynomial Algorithm for Simulating Coalescence with Recombination*, Purdue University Statistics Department Colloquium (December 4, 2003).
5. *Probability Models for Genetic Factors Underlying a Binary Phenotype*, Samuel Lunenfeld Research Institute of Mount Sinai Hospital; Toronto, Canada (March 10, 2000).
6. *A Markov Chain Model of Coalescence with Recombination*, Biophysics Department Seminar; Purdue University (February 9, 2000).
7. *Probability Models for Genetic Factors Underlying a Binary Phenotype*, Canadian Mathematical Society Winter Meeting; Montréal, Canada (December 13, 1999).²
8. *Probability Models for Genetic Factors Underlying a Binary Phenotype*, Statistics Department Colloquium; Purdue University (November 18, 1999).
9. *Properties of Statistical Tests of Neutrality for DNA Polymorphism Data*, Biostatistics / Statistical Genetics Seminar; Purdue University (October 26, 1999).
10. *Genome Data: Finding Trait Genes Using a Dense Marker Map*, 31st Symposium on the Interface: Computing Science and Statistics (June 11, 1999).
11. *Likelihood Ratio Testing with DNA Sequences and Monte-Carlo Integration*, Joint Purdue – Illinois Colloquium, Urbana-Champaign, IL (April 15, 1999).
12. *Likelihood Ratios and DNA Sequences – Testing for Evolution*, GSO, Statistics Department; Purdue University (November 2, 1998).
13. *An Introduction to Statistical Genetics*, Sixth Purdue Symposium on Statistics; Purdue University (June 21, 1998).

²International Meeting, Invited Speaker

14. *Searching for Disease Genes Using Correlated Markers: A Monte-Carlo Approach to Multiple Testing*, Statistics Department Seminar; Purdue University (April 21, 1998).
15. *Likelihood Ratio Testing with DNA Sequences and Monte-Carlo Integration*, Statistics Department Seminar; North Carolina State University (April 1997).
16. *A Markov Chain Model of Recombination*, Statistics Department Colloquium; Purdue University (December 1996).
17. *Choosing an Affection Status Model for Bipolar Affective Disorder*, Genetic Analysis Workshop 10; Pajaro Dunes, CA (October 1996).
18. *Hypothesis Testing with DNA Polymorphism Data*, Postdoctoral Research Symposium, Department of Genetics; North Carolina State University (September 1996).
19. *The Markov Chain Model of Recombination in Coalescence: A Computational Approach*, Canadian Workshop on Computational Biology; University of Western Ontario (January 1996).
20. *The Two-locus Coalescent Model with Recombination*, Eastern Great Lakes Molecular Evolution Meeting; McMaster University (May 1995).

D. Contributed Talks

1. *An Efficient Algorithm For Simulating Coalescence With Recombination*, 36th Symposium on the Interface: Computing Science and Statistics, Baltimore, MD (May 2004).
2. *Multiple Testing Issues for Linked Genetic Markers*, Eastern North American Regional Meeting of the International Biometrics Society, Charlotte, NC (March 2001).