

# CURRICULUM VITAE

Simon TAVARÉ

## BIOGRAPHICAL

email stavare at usc -dot- edu

web <http://www.cmb.usc.edu/people/stavare/Tavare.html>

### Research Interests

Statistics and probability, with particular emphasis on statistical inference in molecular biology, human genetics, population genetics, molecular evolution, paleontology, cancer biology and bioinformatics. Microarray and resequencing data analysis. Fly tracking and behavior. Stochastic computation techniques, coalescent theory, probabilistic combinatorics. Experimental approaches to detecting methylation.

### Education

University of Sheffield, UK    Probability and Statistics    BSc 1974  
University of Sheffield, UK    Probability and Statistics    MSc 1975  
University of Sheffield, UK    Probability and Statistics    PhD 1979

### Professional Experience

2006 –            Senior Group Leader, CR UK Cambridge Research Institute  
2003 –            Professor of Cancer Research (Bioinformatics)  
                      Department of Oncology, University of Cambridge  
2003 –            Professor, Department of Applied Mathematics  
                      and Theoretical Physics, University of Cambridge  
2006 –            Research Professor, Department of Biological Sciences, USC  
2000 – 2006      Professor-at-Large, Keck Graduate Institute of Applied  
                      Life Sciences, Claremont CA  
1989 – 2006      Professor, Department of Biological Sciences, USC  
1999 – 2006      Professor, Department of Preventive Medicine,  
                      Keck School of Medicine, USC  
1989 – 2006      Professor, Department of Mathematics, USC  
1984 – 1989      Associate Professor of Mathematics, University of Utah  
1983 – 1985      Associate Professor of Statistics, Colorado State University  
1981 – 1983      Assistant Professor of Statistics, Colorado State University  
1978 – 1981      Instructor of Mathematics, University of Utah  
1978 – 1979      Research Fellow, Medical Biophysics and Computing,  
                      University of Utah  
1977 – 1978      Lecturer in Probability and Statistics, University of Sheffield.

## Visiting Appointments

Chalmers Jubilee Professor, Chalmers University, Sweden. 8-10/10.  
G. C. Steward Visiting Fellowship in Mathematics, Gonville and Caius College, Cambridge, England. 1-7/2003  
SAMSI, North Carolina. 9-10/2002  
Université Blaise Pascal, Clermont-Ferrand, France. 7/2001  
Institute for Pure and Applied Mathematics, UCLA. 9-12/2000  
Isaac Newton Institute, University of Cambridge, England. 9/1998  
Department of Mathematical Statistics, Chalmers University of Technology and Gothenburg University, Sweden. 5/1998  
Department of Applied Mathematics, University of Zürich, Switzerland. 5-6/1997, 4-7/1996 (Gastprofessor), 8/1993, 12/1992, 5-6/1992.  
Department of Mathematics, Monash University, Australia. 7/1997, 7/1994, 3/1993, 4-5/1990, 7-12/1980.  
Department of Mathematics, Queen Mary and Westfield College, London. 8/1989, 9-12/1988  
Department of Statistical Science, University College, London. 7-8/1988  
Department of Mathematics, University of Southern California. 8-12/1987, 10-11/1984.  
Mathematics Department, Stanford University. 6-8/1982, 6-8/1981, 6-8/1979.

## Honors and Awards

2009	Elected Fellow, Society of Biology (FSB)
2009	Elected Fellow, Academy of Medical Sciences (FMedSci)
2004	Elected Fellow, American Statistical Association
2003 – 2008	Royal Society/Wolfson Research Merit Award
2001	Medallion Lecture, Institute of Mathematical Statistics
1998 –	George and Louise Kawamoto Chair in Biological Sciences, USC
1998	Elected Fellow, American Association for the Advancement of Science
1993	Medallion Lecture, Institute of Mathematical Statistics
1992	Elected Fellow, Institute of Mathematical Statistics

## Professional Qualifications

1993	CStat
2010	CSci

## PUBLICATIONS

### Books

1. Deonier RC, Tavaré S & Waterman MS. *Computational Genome Analysis: An Introduction*, 2005. Springer Verlag, New York. 540pp. ISBN: 0-387-98785-1

2. Arratia R, Barbour AD & Tavaré S. *Logarithmic Combinatorial Structures: A Probabilistic Approach*. EMS Monographs in Mathematics, **1**, 2003. European Mathematical Society Publishing House, Zuerich. 352pp.
3. Donnelly P & Tavaré S. (eds) *Progress in Population Genetics and Human Evolution*. IMA Volumes in Mathematics and its Applications, **87**, 1997. Springer Verlag, Berlin. 330pp.

### Submitted Articles

4. Sottoriva A & Tavaré S. Integrating approximate Bayesian computation with complex agent-based models for cancer research. *COMPSTAT2010*.
5. Speed D & Tavaré S. Sparse partitioning: A method to detect interactions in high dimensional regression problems with binary or tertiary predictors. *Ann Appl Stats*.
6. Smith ML, Dunning MJ, Tavaré S & Lynch AG. The venerable bead: How raw Illumina data allow for the identification and correction of previously unreported spatial phenomena. *BMC Bioinformatics*.
7. Ritchie ME, Forrest MS, Dimas AS, Daelemans C, Dermitzakis ET, Deloukas P & Tavaré S. Data analysis issues for allele-specific expression using Illumina's GoldenGate assay. *BMC Bioinformatics*.
8. Ivakhno S & Tavaré S. CNAnova: a new approach for finding recurrent copy number abnormalities in cancer SNP microarray data. *Bioinformatics*.
9. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, Campino S, Clark TG, Stanier P, Kwiatkowski D, Deloukas P, Dermitzakis ET, Tavaré S, Moore GE & Dunham I. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genomics*.
10. Stirling EJ, Goldman TD, Grover D, Tavaré S & Arbeitman MN. Neurons in the pars intercerebralis regulate circadian rhythm and male courtship behaviors, affecting the timing of social interactions in *Drosophila*. *PLoS Biology*.

## Journal Articles

Reviews are indicated by \*\*

### In Press

11. Clark AM, Goldstein LD, Tevlin M, Tavaré S, Shaham S & Miska EA. The microRNA miR-124 controls gene expression in the sensory nervous system of *Caenorhabditis elegans*. *Nucleic Acids Research*, in press, 2010.
12. †Barbour AD & Tavaré S. Assessing molecular variability in cancer genomes. In *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman*, eds. Bingham N & Goldie C. Cambridge, Cambridge University Press, in press, 2010.
13. \*\* Ivakhno S, Frese K, Tavaré S, Iacobuzio-Donahue C, Tuveson D. Systems biology of pancreatic cancer. In *Systems Biology of Cancer*, ed. Thiagalingam S. Cambridge University Press. In press, 2010.

### 2009

14. Curtis C, Lynch AG, Dunning MJ, Spiteri I, Marioni JC, Hadfield J, Chin S-F, Brenton J, Tavaré S & Caldas C. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics*, **10**: 588, 2009.
15. Barbosa-Morais NL, Dunning MJ, Samarajiwa S, Darot JFJ, Ritchie ME, Lynch AG & Tavaré S. A reannotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Research*, doi:10.1093/nar/gkp942, 2009.
16. Grover G & Tavaré S. Finding behavioral motifs in fly trajectories. *Communications in Information and Systems*, **9**, 283–294, 2009.
17. Avril-Sassen S, Goldstein L, Stingl J, Blenkiron C, Le Quesne J, Spiteri I, Karagavrilidou K, Watson CJ, Tavaré S, Miska EA & Caldas C. Characterization of microRNA expression in post-natal mouse mammary gland development. *BMC Genomics*, **10**: 548, 2009.
18. Waskar M, Landis GN, Shen J, Curtis CN, Abdueva D, Skvortsov D, Tavaré S & Tower J. *Drosophila melanogaster* p53 has developmental stage-specific and sex-specific effects on adult life span indicative of sexual antagonistic pleiotropy. *Aging*, **1**, 1-31, 2009.
19. Grover D, Ford D, Brown C, Hoe N, Erdem E, Tavaré S & Tower J. Hydrogen peroxide stimulates activity and alters behavior in *Drosophila melanogaster*. *PLoS ONE*, **4**: e7580, 2009.

20. Knott S, Viggiani CJ, Aparicio OM & Tavaré S. Strategies for analyzing highly enriched IP-chip datasets. *BMC Bioinformatics*, **10**: 305, 2009.
21. Spyrou C, Stark R, Lynch AG & Tavaré S. BayesPeak: Bayesian analysis of ChIP-seq data. *BMC Bioinformatics*, **10**: 299, 2009.
22. Fernando H, Sewitz S, Darot J, Tavaré S, Huppert JL & Balasubramanian S. Genome-wide analysis of a G-quadruplex-specific single-chain antibody that regulates gene expression. *Nucleic Acids Research*, **37**, 6716–6722, 2009.
23. Ritchie ME, Carvalho BS, Hetrick KN, Tavaré S & Irizarry RA. R/Bioconductor software for Illumina’s Infinium whole-genome genotyping BeadChips. *Bioinformatics*, **25**, 2621–2623, 2009.
24. \*\* Siegmund KD, Marjoram P, Tavaré S & Shibata D. Many colorectal cancers are “flat” clonal expansions. *Cell Cycle*, **8**, 2187–2193, 2009.
25. Tiemann-Boege I, Curtis CN, Shinde DN, Goodman DB, Tavaré S & Arnheim N. Product length, dye choice, and detection chemistry in the bead-emulsion amplification of millions of single DNA molecules in parallel. *Analytical Chemistry*, **81**, 5770–5776, 2009.
26. Knott S, Viggiani CJ, Tavaré S & Aparicio OM. Genome-wide replication profiles indicate an expansive role for Rpd3L in regulating replication initiation timing or efficiency, and reveal genomic loci of Rpd3 function in *Saccharomyces cerevisiae*. *Genes & Dev*, **23**, 1077–1090, 2009.
27. Grover D, Yang J, Ford D, Tavaré S & Tower J. Simultaneous tracking of movement and gene expression in multiple *Drosophila melanogaster* flies using GFP and DsRED fluorescent reporter transgenes. *BMC Research Notes*, **2**: 58, 2009.
28. Wilkinson RD & Tavaré S. Estimating primate divergence times by using conditioned birth-and-death processes. *Theoret Popul Biol*, **75**, 278–285, 2009.
29. Young ARJ, Narita M, Ferreira M, Kirschner K, Sadaie M, Darot JFJ, Tavaré S, Arakawa S, Shimizu S, Watt FM & Narita M. Autophagy mediates the mitotic-senescence transition. *Genes & Dev*, **23**, 798–803, 2009.
30. Siegmund KD, Marjoram P, Woo Y-J, Tavaré S & Shibata D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc Natl Acad Sci USA*, **106**, 4828–4833, 2009.
31. Shen J, Curtis C, Tavaré S & Tower J. A screen of apoptosis and senescence regulatory genes for life span effects when over-expressed in *Drosophila*. *Aging*, **1**, 191–211, 2009.
32. Woo Y-J, Siegmund KD, Tavaré S & Shibata D. Older individuals appear to acquire mitotically older colorectal cancers. *J Pathol*, **217**, 483–488, 2009.

33. Jiang R, Tavaré S & Marjoram P. Population genetic inference from resequencing data. *Genetics*, **181**, 187–197, 2009.
- 2008
34. Grover D, Yang J, Tavaré S & Tower J. Simultaneous tracking of fly movement and expression using GFP. *BMC Biotechnology*, **8**: 93, 2008. (With press release)
  35. Dimas AS, Stranger BE, Beazley C, Finn RD, Ingle CE, Forrest MS, Ritchie ME, Deloukas P, Tavaré S & Dermitzakis ET. Modifier effects between regulatory and protein-coding variation. *PLoS Genetics*, **4** (10): e1000244, 2008.
  36. Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VLJ, Fisher EMC, Tavaré S & Odom DT. Species-specific transcription in mice carrying human chromosome 21. *Science*, **322**, 434–438, 2008.
  37. Hull ML, Rangel Escareno C, Godsland J, Doig J, Johnson C, Butt RP, Phillips S, Smith SK, Tavaré S, Print CG & Charnock-Jones DS. Endometrial-peritoneal interactions during endometriotic lesion establishment. *Am J Pathol*, **173**, 700–715, 2008.
  38. Marioni JC, White M, Tavaré S & Lynch AG. Hidden copy number variation in the HapMap trios. *Proc Natl Acad Sci USA*, **105**, 10067–10072, 2008.
  39. Szyjka SJ, Aparicio JG, Viggiani CJ, Knott S, Xu W, Tavaré S & Aparicio OM. Rad53 regulates replication fork restart after DNA damage in *Saccharomyces cerevisiae*. *Genes & Dev*, **22**, 1902–1920, 2008.
  40. Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, Tomazou EM, Bäckdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJP, Carter NP, Tavaré S & Beck S. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*, **18**, 1518–1829, 2008.
  41. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, Matthews N, Berezikov E, Ketting RF, Tavaré S & Miska EA. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Molecular Cell*, **31**, 79–90, 2008.
  42. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Bäckdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJP, Durbin R, Tavaré S & Beck S. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotechnology*, **26**, 779–785, 2008.

43. Dunning MJ, Ritchie ME, Barbosa-Morais NL, Tavaré S & Lynch AG. Spike-in validation of an Illumina-specific variance-stabilizing transformation. *BMC Research Notes*, **1**: 18, 2008.
  44. Palmer RD, Barbosa-Morais NL, Gooding EL, Muralidhar B, Thornton CM, Pett MR, Roberts I, Schneider DT, Thorne N, Tavaré S, Nicholson JC & Coleman N; On behalf of The Childrens Cancer and Leukaemia Group (CCLG). Pediatric malignant germ cell tumours show characteristic transcriptome profiles. *Cancer Research*, **68**: 4239–4247, 2008.
  45. Chu M, Siegmund KD, Hao Q-L, Crooks GM, Tavaré S & Shibata D. Inferring relative numbers of human leukocyte genome replications. *British Journal of Haematology*, **141**: 862–871, 2008.
  46. Grover D, Tower J & Tavaré S. O fly, where art thou? *J R Soc Interface*, **5**: 1181–1191, 2008.
  47. Dunning MJ, Barbosa-Morais NL, Lynch AG, Tavaré S & Ritchie ME. Statistical issues in the analysis of Illumina data. *BMC Bioinformatics*, **9**: 85, 2008.
- 2007
48. Curtis C, Landis GN, Folk D, Wehr NB, Hoe N, Waskar M, Abdueva D, Skvortsov D, Ford D, Luu A, Badrinath A, Levine RL, Bradley TJ, Tavaré S & Tower J. Transcriptional profiling of MnSOD-mediated lifespan extension in *Drosophila* reveals a species-general network of aging and metabolic genes. *Genome Biology*, **8(12)**: R262, 2007.
  49. Marioni JC, Thorne NP, Valsesia A, Fitzgerald T, Redon R, Fiegler H, Andrews DT, Stranger BE, Lynch AG, Dermitzakis ET, Carter NP, Tavaré S & Hurles ME. Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biology*, **8(10)**: R228, 2007.
  50. Conboy CM, Spyrou C, Thorne NP, Wade EJ, Barbosa-Morais N, Bhattacharjee A, Young RA, Tavaré S, Lees JA & Odom DT. Cell cycle genes are the evolutionarily conserved targets of mammalian DREAM member E2F4. *PLoS ONE*, **2**: e1061, 2007.
  51. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, Costa J, Pinder SE, van de Wiel MA, Green AR, Ellis IO, Porter PL, Tavaré S, Brenton JD, Ylstra B & Caldas C. High-resolution array-CGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biology*, **8(10)**: R215, 2007.
  52. Blenkiron C, Goldstein LD, Thorne NP, Spiteri MI, Chin SF, Dunning M, Barbosa-Morais NL, Tavaré S, Caldas C & Miska EA. MicroRNA expression profiling of human

- breast cancer identifies new markers of tumour subtype. *Genome Biology*, **8** (10): R214, 2007.
53. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle C, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P & Dermitzakis ET. Population genomics of human gene expression. *Nat Genet*, **39**, 1217–1224, 2007.
  54. \*\* Shibata D & Tavaré S. Stem cell chronicles: Autobiographies within genomes. *Stem Cell Reviews*, **3**, 94–103, 2007.
  55. Martin RD, Soligo C & Tavaré S. Primate origins: Implications of a Cretaceous ancestry. *Folia Primatol*, **78**, 277–296, 2007.
  56. Chu MW, Siegmund KD, Eckstam CL, Kim JY, Yang AS, Kanel GC, Tavaré S & Shibata D. Lack of increases in methylation at three CpG-rich genomic loci in non-mitotic adult tissues during aging. *BMC Medical Genetics*, **8**:50, 2007.
  57. Lynch AG, Marioni JC & Tavaré S. Numbers of CNVs and false negative rates will be underestimated if we do not account for the dependence between repeated experiments. *Amer J Hum Genet*, **81**, 418–420, 2007.
  58. Dunning MJ, Smith ML, Ritchie ME & Tavaré S. *beadarray*: R classes and methods for Illumina bead-based data. *Bioinformatics*, **23**, 2183–2184, 2007.
  59. Skvortsov D, Abdueva D, Curtis CN, Schaub B & Tavaré S. Explaining differences in saturation levels for Affymetrix GeneChip arrays. *Nucleic Acids Research*, **35**, 4154–4163, 2007.
  60. Skvortsov D, Abdueva D, Stitzer M, Finkel SA & Tavaré S. Using expression arrays for copy number detection: an example from *E. coli*. *BMC Bioinformatics*, **8**:203, 2007.
  61. Chin S-F, Wang Y, Thorne NP, Teschendorff AE, Pinder SE, Vias M, Naderi A, Roberts I, Barbosa-Morais NL, Garcia M, Iyer NG, Kranjac T, Robertson J, Ruffalo T, Aparicio SFR, Tavaré S, Ellis I, Brenton JD, Caldas C. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene*, **26**, 1959–1970, 2007.
  62. Nicolas P, Kim K-M, Shibata D & Tavaré S. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS Computational Biology*, **3**, 364–374, 2007.
  63. Stranger BE, Forrest MS, Dunning M, Thorne NP, Ingle CE, Beazley C, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME & Dermitzakis ET. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853, 2007.

2006

64. \*\* Dunning M, Smith M, Thorne NP, Tavaré S. beadarray: an R package to analyse Illumina BeadArrays. *R News*, **6**(5), 17–23, 2006.
65. Xu W, Aparicio JG, Aparicio OM & Tavaré S. Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and precise identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics*, **7**: 276, 2006.
66. Ibrahim AEK, Thorne NP, Baird K, Barbosa-Morais NL, Tavaré S, Collins VP, Wylie AH, Arends MJ & Brenton JD. Mmass: an optimised array-based method for assessing CpG island methylation. *Nucleic Acids Research*, **34**: e136, 2006.
67. Abdueva D, Skvortsov D & Tavaré S. Non-linear analysis of GeneChip arrays. *Nucleic Acids Research*, **34**, e105, 2006.
68. \*\* Marjoram P & Tavaré S. Modern computational approaches for analysing molecular-genetic variation data. *Nat Rev Genet*, **7**, 759–770, 2006.
69. Jiang R, Marjoram P, Borevitz JO & Tavaré S. Inferring population parameters from single-feature polymorphism data. *Genetics*, **173**, 2257–2267, 2006.
70. \*\* Shibata D & Tavaré S. Counting divisions in a human somatic cell tree: how, what and why. *Cell Cycle*, **5**, 610–614, 2006.
71. Millman K, Black CM, Stamm WE, Jones RB, Hook EW, Martin DH, Bolan G, Tavaré S & Dean D. Population-based genetic epidemiologic analysis of *Chlamydia trachomatis* serotypes and lack of associations between *ompA* polymorphisms and clinical phenotypes. *Microbes and Infection*, **8**, 604–611, 2006.
72. Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C & Nordborg M. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Mol Biol Evol*, **23**, 1217–1231, 2006.
73. Marioni JC, Thorne NP & Tavaré S. BioHMM: a heterogeneous hidden Markov model for segmenting array CGH data. *Bioinformatics*, **22**, 1144–1146, 2006.
74. Dunning M, Thorne NP, Camilier I, Smith ML & Tavaré S. Quality control and low-level statistical analysis of Illumina BeadArrays. *REVSTAT*, **4**, 1–30, 2006.
75. Arratia R, Barbour AD & Tavaré S. A tale of three couplings: Poisson-Dirichlet and GEM approximations for random permutations. *Combin Probab Comput*, **15**, 31–62, 2006.
76. Kim JY, Tavaré S & Shibata D. Human hair genealogies and stem cell latency. *BMC Biology*, **4**:2, 2006.

2005

77. Stranger BE, Forrest MS, Clark AG, Minichiello M, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P & Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet*, **1**, 695–704, 2005.
78. Kim JY, Tavaré S & Shibata D. Counting human somatic cell replications: Methylation mirrors human endometrial stem cell divisions. *Proc Natl Acad Sci USA*, **102**, 17739–17744, 2005.
79. Clark RM, Tavaré S & Doebley J. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol Biol Evol*, **22**, 2304–2312, 2005.
80. Calabrese P, Mecklin JP, Järvinen HJ, Aaltonen LA, Tavaré S & Shibata D. Numbers of mutations to different types of colorectal cancer. *BMC Cancer*. **5**:126, 2005.
81. Kim JY, Siegmund KD, Tavaré S & Shibata D. Age-related human small intestine methylation: evidence for stem cell niches. *BMC Medicine*. **3**:10, 2005.
82. Arratia R, Barbour AD & Tavaré S. A probabilistic approach to analytic arithmetic on algebraic function fields. *Math Proc Camb Phil Soc*, **139**, 1–26, 2005.
83. Innan H, Zhang K, Marjoram P, Tavaré S & Rosenberg NA. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics*, **169**, 1763–1777, 2005.

2004

84. Tavaré S. Ancestral inference in population genetics. In *Lectures on Probability Theory and Statistics. Ecole d'Etés de Probabilité de Saint-Flour XXXI – 2001*. Ed. J. Picard. Lecture Notes in Mathematics, **1837**, 1–188, 2004. Springer Verlag, New York.
85. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, Tavaré S & Tower J. Similar gene expression patterns characterize aging and oxidative stress in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*, **101**, 7663–7668, 2004.
86. Millman KL, Black CM, Johnson R, Stamm W, Jones R, Hook E, Martin D, Bolan G, Tavaré S, & Dean D. Population-based genetic and evolutionary analysis of *Chlamydia trachomatis* urogenital strain variation in the United States. *J Bacteriol*, **186**, 2457–2465, 2004.
87. Calabrese P, Tavaré S & Shibata D. Pre-tumor progression: clonal evolution of human stem cell populations. *Am J Pathol*, **164**, 1337–1346, 2004.

88. Calabrese P, Tsao J-L, Yatabe Y, Salovaara R, Mecklin J-P, Järvinen HJ, Aaltonen LA, Tavaré S & Shibata D. Colorectal pre-tumor progression before and after loss of DNA mismatch repair. *Am. J. Pathol.*, **164**, 1447–1453, 2004.

89. Kim K-M, Calabrese P, Tavaré S & Shibata D. Enhanced stem cell survival in familial adenomatous polyposis. *Am. J. Pathol.*, **164**, 1369–1377, 2004.

2003

90. Marjoram P, Molitor J, Plagnol V & Tavaré S. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, **100**, 15324–15328, 2003.

2002

91. Tavaré S, Marshall CR, Will O, Soligo C & Martin RD. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, **416**, 726–729, 2002.

92. \*\* Nordborg M & Tavaré S. Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, **18**, 83–90, 2002.

2001

93. Millman KL, Tavaré S & Dean D. Recombination in the *ompA* gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance and persistence of the organism. *Journal of Bacteriology*, **183**, 5997–6008, 2001.

94. Yatabe Y, Tavaré S & Shibata D. Investigating stem cells in human colon by using methylation patterns. *Proc. Natl. Acad. Sci. USA*, **98**, 10839–10844, 2001. (With editorial.)

95. Markovtsova L, Marjoram P & Tavaré S. On a test of Depaulis and Veuille. *Mol. Biol. Evol.*, **18**, 1132–1133, 2001.

96. Tavaré S, Adams DC, Fedrigo O & Naylor GJP. A model for phylogenetic inference using structural and chemical covariates. *Pacific Symposium on Biocomputing*, **6**, 215–225, 2001.

2000

97. Arratia R, Barbour AD & Tavaré S. Limits of logarithmic combinatorial structures. *Ann. Probab.*, **28**, 1620–1644, 2000.

98. Markovtsova L, Marjoram P & Tavaré S. The effects of rate variation on ancestral inference in the coalescent. *Genetics*, **156**, 1427–1436, 2000.

99. Markovtsova L, Marjoram P & Tavaré S. The age of a unique event polymorphism. *Genetics*, **156**, 401–409, 2000.
100. Arratia R, Barbour AD & Tavaré S. The number of components in a logarithmic combinatorial structure. *Ann. Appl. Prob.*, **10**, 331–361, 2000.
101. Schbath S, Bossard N & Tavaré S. The effect of non-homogeneous clone length distribution on the progress of an STS mapping project. *J. Comp. Biol.*, **7**, 47–57, 2000.
102. Tsao JL, Yatabe Y, Salovaara R, Jarvinen HJ, Mecklin JP, Aaltonen LA, Tavaré S & Shibata D. Genetic reconstruction of individual colorectal tumor histories. *Proc. Natl. Acad. Sci. USA*, **97**, 1236–1241, 2000.

1999

103. Tavaré S. Random trees in molecular genetics. *Bull. I.S.I.*, **52**, 269–272, 1999.
104. Jaruzelska J, Zietkiewicz E, Batzer M, Cole D, Moisan J-P, Scozzari R, Tavaré S & Labuda D. Spatial and temporal distribution of the neutral polymorphisms in the last ZFX intron: analysis of the haplotype structure and genealogy. *Genetics*, **152**, 1091–1101, 1999.
105. Fan F, Liu C, Tavaré S & Arnheim N. Polymorphisms in the human DNA repair gene XPF. *Mutation Research Genomics*, **406**, 115–120, 1999.
106. Leeflang E, Tavaré S, Marjoram P, Neal COS, Srinidhi J, MacFarlane H, MacDonald ME, Gusella JF, de Young M, Wexler N & Arnheim N. Analysis of germline mutation spectra at the Huntington Disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.*, **8**, 173–183, 1999.
107. Tsao JL, Tavaré S, Salovaara R, Jass JR, Aaltonen LA & Shibata D. Colorectal adenoma and cancer divergence: evidence of multi-lineage progression. *Am. J. Pathol.*, **154**, 1815–1824, 1999.
108. Griffiths RC & Tavaré S. The ages of mutations in gene trees. *Ann. Appl. Prob.*, **9**, 567–590, 1999.
109. Arratia R, Barbour AD & Tavaré S. The Poisson-Dirichlet distribution and the scale-invariant Poisson process. *Combin. Probab. Comput.*, **8**, 407–416, 1999.
110. Arratia R, Barbour AD & Tavaré S. On Poisson-Dirichlet limits for random decomposable combinatorial structures. *Combin. Probab. Comput.*, **8**, 193–208, 1999.

1998

111. Griffiths RC & Tavaré S. The age of a mutation in a general coalescent tree. *Stochastic Models*, **14**, 273–295, 1998.

1997

112. Arratia R, Barbour AD & Tavaré S. Random combinatorial structures and prime factorizations. *Notices of the American Mathematical Society*, **44**, 903–910, 1997.
113. Tavaré S, Balding DJ, Griffiths RC & Donnelly P. Inferring coalescence times for molecular sequence data. *Genetics*, **145**, 505–518, 1997.

1996

114. Shenkar R, Navidi W, Tavaré S, Dang MH, Chomyn A, Attardi G, Cortopassi G & Arnheim N. The mutation rate of the human mitochondrial DNA deletion mtDNA<sup>4977</sup>. *Amer. J. Hum. Gen.*, **59**, 772–780, 1996. (With editorial.)
115. Donnelly P, Tavaré S, Balding DJ & Griffiths RC. Estimating the age of the common ancestor of men from the ZFY intron. *Science*, **272**, 1357–1359, 1996.
116. Griffiths RC & Tavaré S. Monte Carlo inference methods in population genetics. *Mathl. Comput. Modelling*, **23**, 141–158, 1996.

1995

117. Donnelly P & Tavaré S. Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.*, **29**, 401–421, 1995.
118. Leeflang EP, Zhang L, Tavaré S, Hubert R, Srinidhi J, MacDonald ME, Myers RM, de Young M, Wexler NS, Gusella JF & Arnheim N. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum. Molec. Genet.*, **4**, 1519–1526, 1995.
119. Arratia R, Stark D & Tavaré S. Total variation asymptotics for Poisson process approximations of logarithmic combinatorial assemblies. *Ann. Probab.*, **23**, 1347–1388, 1995.
120. Joyce P & Tavaré S. The distribution of rare alleles. *J. Math. Biol.*, **33**, 602–618, 1995.
121. Griffiths RC & Tavaré S. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math. Biosci.*, **127**, 77–98, 1995.

1994

122. Griffiths RC & Tavaré S. Ancestral inference in population genetics. *Statistical Science*, **9**, 307–319, 1994.
123. Griffiths RC & Tavaré S. Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. Lond. B*, **344**, 403–410, 1994.

124. Barbour AD & Tavaré S. A rate for the Erdős-Turán law. *Combin. Probab. Comput.*, **3**, 167–176, 1994.
125. Griffiths RC & Tavaré S. Simulating probability distributions in the coalescent. *Theor. Popn. Biol.*, **46**, 131–159, 1994.
126. Arratia R & Tavaré S. Independent process approximations for random combinatorial structures. *Adv. Math.*, **104**, 90–154, 1994.
127. Raftery AE & Tavaré S. Estimation and modelling repeated patterns in high order Markov chains with the mixture transition distribution (MTD) model. *Applied Statistics*, **43**, 179–199, 1994.
- [1993](#)
128. Arratia R, Barbour AD & Tavaré S. On random polynomials over a finite field. *Math. Proc. Camb. Phil. Soc.*, **114**, 347–368, 1993.
- [1992](#)
129. Lundstrom R, Tavaré S & Ward RH. Modelling the evolution of the human mitochondrial genome. *Math. Biosci.*, **112**, 319–336, 1992.
130. Lundstrom R, Tavaré S & Ward RH. Estimating substitution rates from molecular data using the coalescent. *Proc. Natl. Acad. Sci., USA*, **89**, 5961–5965, 1992.
131. Arratia R & Tavaré S. Limit theorems for combinatorial structures via discrete process approximations. *Rand. Struct. Alg.*, **3**, 321–345, 1992.
132. Arratia R, Barbour AD & Tavaré S. Poisson process approximations for the Ewens Sampling Formula. *Ann. Appl. Probab.*, **2**, 519–535, 1992.
133. Arratia R & Tavaré S. The cycle structure of random permutations. *Ann. Probab.*, **20**, 1567–1591, 1992.
134. Joyce PJ & Tavaré S. A convergence theorem for symmetric functionals of random partitions. *J. Appl. Prob.*, **29**, 280–290, 1992.
- [1991](#)
135. Donnelly P, Kurtz TG & Tavaré S. On the functional central limit theorem for the Ewens sampling formula. *Ann. Appl. Probab.*, **1**, 539–545, 1991.
136. Arratia R, Lander ES, Tavaré S & Waterman MS. Genomic mapping by anchoring random clones: a mathematical analysis. *Genomics*, **11**, 806–827, 1991.

1990

137. Joyce PJ & Tavaré S. Random permutations and neutral evolution models. *Stoch. Proc. and Applns.*, **36**, 245–262, 1990.

1989

138. Tavaré S, Ewens WJ & Joyce P. Is knowing the age-order of alleles useful in testing for selective neutrality? *Genetics*, **122**, 705–711, 1989.
139. Tavaré S & Song B. Codon preference and primary sequence structure in protein coding regions. *Bull. Math. Biol.*, **51**, 95–115, 1989.

1987

140. Joyce PJ & Tavaré S. Cycles, permutations and the structure of the Yule process with immigration. *Stoch. Proc. and Applns.*, **25**, 309–314, 1987.
141. Donnelly P & Tavaré S. The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.*, **25**, 381–391, 1987.
142. Tavaré S. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.*, **25**, 161–168, 1987.

1986

143. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86, 1986.
144. Donnelly P & Tavaré S. The ages of alleles and a coalescent. *Adv. Appl. Prob.*, **18**, 1–19, 1986. Correction:**18**, 1023.

1984

145. Tavaré S. Lines of descent and genealogical processes, and their application in population genetics models. *Theoret. Popn. Biol.*, **26**, 119–164, 1984.
146. Saunders IW, Tavaré S & Watterson GA. On the genealogy of nested subsamples from a haploid population. *Adv. Appl. Prob.*, **16**, 471–491, 1984.

1983

147. Karlin S, Ghandhour G, Ost F, Tavaré S & Korn LJ. New approaches for computer analysis of nucleic acid sequences. *Proc. Natl. Acad. Sci. USA*, **80**, 5660–5664, 1983.
148. Ethier SN & Tavaré S. The proportional bettor's return on investment. *J. Appl. Prob.*, **20**, 563–573, 1983.

149. Seneta E & Tavaré S. Some stochastic models for plasmid copy number. *Theoret. Popn. Biol.*, **23**, 241–256, 1983.
150. Seneta E & Tavaré S. A note on models using the branching process with immigration stopped at zero. *J. Appl. Prob.*, **20**, 11–18, 1983.
151. Tavaré S. Serial dependence in contingency tables. *J. Royal Statist. Soc. B*, **45**, 100–106, 1983.
152. Tavaré S & Altham PME. Serial dependence of observations leading to contingency tables, with corrections to  $\chi^2$  statistics. *Biometrika*, **70**, 139–144, 1983.
153. Karlin S & Tavaré S. A class of diffusion stochastic processes with killing arising in population genetics. *SIAM J. Appl. Math.*, **43**, 31–41, 1983.
- 1982
154. Karlin S & Tavaré S. Linear birth and death processes with killing. *J. Appl. Prob.*, **19**, 477–487, 1982.
155. Karlin S & Tavaré S. Detecting particular genotypes in populations under non-random mating. *Math. Biosci.*, **59**, 57–75, 1982.
156. Karlin S & Tavaré S. A diffusion with killing: the time to formation of recurrent deleterious mutant genes. *Stoch. Proc. and Applns.*, **13**, 249–261, 1982.
- 1981
157. Pakes AG & Tavaré S. Comments on the age distribution of Markov processes. *Adv. Appl. Prob.*, **13**, 681–703, 1981.
158. Karlin S & Tavaré S. The detection of particular genotypes in finite populations. II. The effects of partial penetrance and family structure. *Theoret. Popn. Biol.*, **19**, 215–229, 1981.
159. Karlin S & Tavaré S. The detection of particular genotypes in finite populations. I. Natural selection effects. *Theoret. Popn. Biol.*, **19**, 187–214, 1981.
160. Karlin S & Tavaré S. The detection of a recessive visible gene in finite populations. *Genetical Research*, **37**, 22–46, 1981.
- 1980
161. Tavaré S. Time reversal and age distributions. I. Discrete-time Markov chains. *J. Appl. Prob.*, **17**, 33–46, 1980.

1979

162. Tavaré S. Dual diffusions, killed diffusions and the age distribution problem in population genetics. *Theoret. Popn. Biol.*, **16**, 253–266, 1979.
163. Tavaré S. A note on finite continuous-time Markov chains. *Biometrics*, **35**, 831–834, 1979.
164. Tavaré S. Sojourn times for conditioned Markov chains in genetics. *Theoret. Popn. Biol.*, **15**, 108–112, 1979.
165. Oakley JR & Tavaré S. Another approach to the assessment of growth in early infancy. *Developmental Medicine and Child Neurology*, **21**, 186–193, 1979.

**Book Chapters** †denotes refereed article

166. \*\* Lynch AG, Smith ML, Dunning MJ, Cairns JM, Barbosa-Morais NL & Tavaré S. *beadarray*, BASH and HULK – tools to increase the value of Illumina BeadArray experiments. In *Statistical Tools for Challenges in Bioinformatics*, eds. Gusnanto A, Mardia KV, & Fallaize CJ. Leeds, Leeds University Press, pp. 33–37, 2009.
167. †\*\* Thorne NP, Marioni JC, Rakyan V, Ibrahim AEK, Massie C, Curtis C, Brenton JD, Murrell A & Tavaré S. DNA methylation arrays: Methods and analysis. In *Microarray Innovations: Technology and Experimentation in Drug Discovery and Biomedical Research*, ed. G. Hardiman. CRC Press/Taylor and Francis, Ch. 13, pp. 175–206, 2009.
168. Lynch AG, Curtis CN & Tavaré S. Correcting for probe-design in the analysis of gene-expression microarrays. In *Systems Biology & Statistical Bioinformatics*, eds. Barber S, Baxter PD & Mardia KV. Leeds, Leeds University Press, pp. 83–86, 2007.
169. Tavaré S, Nicolas P & Shibata D. Stochastic models for the evolution of stem cells in colon crypts. In *Systems Biology & Statistical Bioinformatics*, eds. Barber S, Baxter PD & Mardia KV. Leeds, Leeds University Press, pp. 61–64, 2007.
170. †Journée M, Teschendorff AE, Absil P-A, Tavaré S & Sepulchre R. Geometric optimization methods for the analysis of gene expression data. In *Principal Manifolds for Data Visualization and Dimension Reduction*, eds. A. Gorban, B. Kegl, D. Winsch & A. Zinovyev. LNCSE, Springer, pp. 271–292, 2007.
171. †Soligo C, Will O, Tavaré S, Marshall CR & Martin RD. New light on the dates of primate origins and divergence. In *Primate Origins: Adaptations and Evolution*. Eds. MJ Ravosa & M Dagosto. Springer Verlag, New York, pp. 29–49, 2007. ISBN: 0-387-30335-9

172. †Tavaré S. Ancestral inference for branching processes. In *Branching Processes in Biology: Variation, Growth, Extinction*, eds. P Haccou, P Jagers & V Vatutin. Cambridge University Press, pp. 208–217, 2005.
173. †Plagnol V & Tavaré S. Approximate Bayesian computation and MCMC. In *Monte Carlo and Quasi-Monte Carlo Methods 2002*, ed. H Niederreiter, Springer-Verlag, pp. 99–114, 2004.
174. †Navidi WC, Tavaré S & Arnheim N. The roles of mutation rate and selective pressure on observed levels of the human mitochondrial DNA deletion mtDNA<sup>4977</sup>. In *Science and Statistics: A Festschrift for Terry Speed*, ed. DR Goldstein, IMS Lecture Notes – Monograph Series, Volume 40, pp. 247–258, 2003.
175. †\*\* Tavaré S. Coalescent theory. In *Nature Encyclopedia of the Human Genome*, Vol. 1, ed. Cooper DN, pp. 836–845, 2003. Nature Publishing Group, London.
176. †Griffiths RC & Tavaré S. The genealogy of a neutral mutation. In *Highly Structured Stochastic Systems*. Eds. Green P, Hjort N and Richardson S, pp. 393–412, 2003. Oxford Statistical Science Series, Vol. 27, Oxford University Press.
177. \*\* Neuhauser C & Tavaré S. The coalescent. *Encyclopedia of Genetics*, Vol. I. Eds. Brenner S and Miller J, pp. 392–397, 2001. Academic Press, New York.
178. Barbour AD & Tavaré S. A rate for the Erdős-Turán law. In *Combinatorics, geometry and probability. A tribute to Paul Erdős*. Eds. Bollobás B and Thomason A, pp. 71–80, 1998. Cambridge University Press. (A reprinting of article 124 above)
179. Leeflang EP, Tavaré S, Marjoram P, Grewal R, Neal COS & Arnheim N. Human germline mutation analysis by single genome PCR: application to dynamic mutations. Chapter 37 in *Genetic instabilities and hereditary neurological diseases*. Eds. Wells RD and Warren ST, pp. 543–560, 1998. Academic Press, New York.
180. †\*\* Ewens WJ & Tavaré S. The Ewens Sampling Formula. In *Encyclopedia of Statistical Science*, Vol. 2 update. Eds. Kotz S, Read CB and Banks DL, pp. 230–234, 1998. Wiley, New York.
181. †Tavaré S. Ancestral inference from DNA sequence data. Chapter 5 in *Case Studies in Mathematical Modeling: Ecology, Physiology, and Cell Biology*. Eds. Othmer HG, Adler FR, Lewis MA and Dallon J, pp. 81–96, 1997. Prentice-Hall.
182. †\*\* Tavaré S & Ewens WJ. Multivariate Ewens distribution. Chapter 41 in *Discrete Multivariate Distributions*. Eds. Johnson NS, Kotz S and Balakrishnan N, pp. 232–246, 1997. Wiley, New York.

183. †Griffiths RC & Tavaré S. Computational methods for the coalescent. Chapter 10 in *Progress in Population Genetics and Human Evolution*. Eds. P. Donnelly P and Tavaré S. IMA Volumes in Mathematics and its Applications, **87**, 165–182, 1997. Springer Verlag, Berlin.
184. †Tavaré S. Calibrating the clock: using stochastic processes to measure the rate of evolution. Chapter 5 in *Calculating the Secrets of Life*, Eds. Lander ES and Waterman MS, pp. 114–152, 1995. National Academy Press, Washington.
185. Tavaré S. The genealogy of the birth, death and immigration process. Chapter 3 in *Mathematical Evolutionary Theory*. Ed. Feldman MW, pp. 41–56, 1989. Princeton University Press.
186. †Tavaré S & Giddings BW. Some statistical aspects of the primary structure of nucleotide sequences. *Mathematical Methods for DNA Sequences*. Ed. Waterman MS, pp. 117–132, 1988. CRC Press.
187. Donnelly P & Tavaré S. A genealogical description of the infinitely-many neutral alleles model. *Stochastic Methods in Biology*. Eds. Kimura M, Kallianpur G, Hida T. Lecture Notes in Biomathematics, **70**, 27–35, 1987. Springer-Verlag.
188. †Williamson JA, Bishop DT, Tavaré CJ & Tavaré S. On the independence structure that exists within a pedigree with an application to testing for a major gene. In *Progress in Clinical and Biological Research*. Eds. Sing CF and Skolnick MH. **32**, 271–276, 1979. AR Liss, New York.

### Book Reviews

189. Arratia R & Tavaré S. *Probability approximations via the Poisson clumping heuristic* by Aldous D, *Poisson approximation* by Barbour AD, Holst L and Janson S. *Ann. Probab.*, **21**, 2269–2279, 1993.
190. Tavaré S. *Mathematical Methods of Applied Probability*, by Hunter JJ. *J. A. S. A.*, **80**, 1071–1072, 1985.
191. Tavaré S. *Genealogical and Genetic Structure*, by Cannings C and Thompson EA. *J. A. S. A.*, **79**, 743–744, 1984.

## Abstracts, Conference Reports and Other Manuscripts

192. Papatheodorou I, Crichton C, Morris L, Maccallum P, METABRIC Group, Davies J, Brenton JD & Caldas C. A metadata approach for clinical data management in translational genomics studies in breast cancer. *BMC Medical Genomics*, **2**: 66, 2009.
193. Marjoram P, Markovtsova L, & Tavaré S. “I see dead people:” gene mapping via ancestral inference. Draft for *Genetic Analysis Workshop 12*. Available at [www.cmb.usc.edu/people/stavare/allstpapers.html](http://www.cmb.usc.edu/people/stavare/allstpapers.html)
194. Marjoram P, Markovtsova L & Tavaré S. Estimation in ancestral recombination graphs using Markov chain Monte Carlo. Research Report, October 2000. Available at [www.cmb.usc.edu/people/stavare/allstpapers.html](http://www.cmb.usc.edu/people/stavare/allstpapers.html)
195. Markovtsova L, Marjoram P & Tavaré S. Discussion of ‘Inference in molecular population genetics’ by Stephens and Donnelly. *J. Royal Statist. Soc. B*, **62**, 647, 2000.
196. Martin RD, Soligo C, Tavaré S, Will O & Marshall CR. New light on the dates of primate origins and divergence. *Folia Primatol.*, **71**, 358–359, 2000.
197. Martin RD, Soligo C, Tavaré S, Will O & Marshall CR. Early primates and the effects of a fragmentary fossil record on dating evolutionary divergences. *Abstracts of the International Symposium on the Evolution of Vertebrates*. University of Lund, Sweden, 1999.
198. Marshall CR, Tavaré S, Will O, Soligo C & Martin RD. Taking into account the temporal ranges of unobserved species: reconciliation of fossil and molecular clock estimates of primate origins? *Geological Society of America Abstracts with Programs*, **199830 (7)**, A-326, 1998.
199. Tavaré S & Feng Y. Reconstructing phylogenetic trees when sites are dependent. Pp. 55–57 in *Proceedings of Phylogeny Workshop*, DIMACS Technical Report 95-48, ed. S. Tavaré, 1995. [See <http://dimacs.rutgers.edu/Workshops/Phylogeny/>]
200. Tavaré S. International Conference on Random Mappings, Partitions, and Permutations. Report, *Adv. Appl. Prob.*, **24**, 761–777, 1992.
201. Tavaré S. Mathematical models in population genetics. *Composite Function*, 1991. (A reprinting of article 209 above)
202. Tavaré S & Lundstrom R. Comments on ‘Sampling strategies for distances between DNA sequences’, by Weir and Basten. *Biometrics*, **46**, 577–582, 1990.
203. Tavaré S & Doerge RW. An application of computer algebra to likelihood analysis. Manuscript, 19pp, 1988.

204. Donnelly P & Tavaré S. The genealogy of the infinite alleles model. In *Proceedings, First World Congress of the Bernoulli Society, 1986*. VNU Sci. Press, Utrecht, 1987, pp. 701–704.
205. Tavaré S. A genealogical view of some stochastic models in population genetics. *Stoch. Proc. Applns.*, **19**, 10, 1985.
206. Tavaré S & Janzen T. On estimating substitution rates from pairs of homologous nucleotide sequences. Manuscript, 28pp, 1985.
207. Tavaré S. The estimation of substitution rates and divergence times from DNA sequence data. *Third Rocky Mountain Regional Conference on Medical Applications of Statistics*, Preprint Volume, 89–96, 1985.
208. Seneta E & Tavaré S. Stochastic models for plasmid copy number. *First Rocky Mountain Regional Conference on Medical Applications of Statistics*, Preprint Volume, 27–33, 1982.
209. Tavaré S. Mathematical models in population genetics. *Function*, **4**, 26–32, 1980.
210. Tavaré S. Time reversion and age distributions. *Adv. Appl. Prob.*, **11**, 10–12, 1979.
211. Tavaré S. Conference on ecological statistics - a summary report. *J. Royal Statist. Soc. C*, **27**, 344–345, 1978.
212. Tavaré S. Age distributions for Markov chains in genetics. *Adv. Appl. Prob.*, **10**, 17–19, 1978.
213. Tavaré S. Sojourn times for conditional Markov chains in genetics. *Adv. Appl. Prob.*, **8**, 645–647, 1976.

#### ACTIVE GRANT SUPPORT

National Institutes of Health P50 HG02790, Center of Excellence in Genome Science. *Genomic analysis of the genotype-phenotype map*. PI. 9/09–8/14.

National Institutes of Health R01 MH084678. *Statistical methods for relating sequence data to phenotype*. Co-investigator (P. Marjoram PI). 9/08–6/11.

## PROFESSIONAL SERVICE

### **Advisory Boards**

Advisory Board, PUMPKIN (Membrane pumps in cells and disease, Aarhus University) (2007 – )

Advisory Board, Gene Ontology Consortium (2006 – )

Advisory Board, BioConductor (Chair; 2006 – )

Advisory Board, Gothenburg Mathematical Modelling Centre, Sweden. (2006 – )

Advisory Board, Cornell Research Training Group in Probability (2005 – )

Steering Committee for Statistics, Eurandom, Holland. (2004 – )

Science Advisory Board, BioDiscovery, Inc. (2003 – )

Science Advisory Board, Institute for Pure and Applied Mathematics (IPAM), UCLA (2002 – 2008)

Senior Advisory Board, NESCent, North Carolina. (2004 – 2007)

International Advisory Board, Swegene, Sweden (2001 – 2006)

Scientific Advisory Panel, The Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, New Zealand. (2003 – 2005)

Science Advisory Board, Amersham plc. (2001 – 2004)

Scientific Advisory Board, Chalmers University of Technology Bioscience Initiative (1999 – 2001)

Scientific Advisory Board, GeneData AG (1997 – 2002)

### **Society Membership**

American Association for the Advancement of Science (1991 – )

American Statistical Association (1992 – )

Institute of Mathematical Statistics (1978 – )

Member, Committee on Fellows, 1996–99.

Royal Statistical Society (1974 –)

Genetics Society of America (1990 –)

### **Editorial and Refereeing Services**

Member, Biological Sciences Funding Committee, CRUK. (2008 – )

Cancer Prevention Research (Editorial Board, 2008 – )

MathematicS In Action (Editorial Board, 2008 – )

Annals of Applied Statistics (Associate Editor, 2007 – )

Statistics Surveys (IMS Associate Editor, 2007 – )

Evolutionary Bioinformatics (Editorial Board, 2005 – )

Human Genomics (Editorial Board, 2003 – )

Annals of Human Genetics (Associate Editor, 2003 – )

Journal of Computational Biology (Associate Editor, 1997 – )

Journal of the Royal Society Interface (Editorial Board, 2003 – 2009)

Genetics (Associate Editor, 1997 – 2003)

Journal of Mathematical Biology (Editorial Board, 2000 – 2003)

Bernoulli (Associate Editor, 2000 – 2003)

Annals of Probability (Associate Editor, 2000 – 2002)

Combinatorics, Probability and Computing (Editorial Board, 1996 – 2001)

Annals of Applied Probability (Associate Editor, 1993 – 1999)

Stochastic Models (Editorial Board, 1991 – 2001)

Mathematical Biosciences (Editorial Board, 1990 – 2001)

Stochastic Processes and their Applications (Associate Editor, 1989 –1993)

Theoretical Population Biology (Associate Editor, 1985 –1987)

Referee for 30+ journals in statistics, probability, combinatorics and biology

Advisory Board, Kluwer Series in Computational Biology (1998 – present)

Referee for numerous grant proposals (e.g. CRUK, BBSRC, EPSRC, NSF, NIH, NSA, Wellcome Trust, Keck, Swiss Nationalfond)

### **Postdoctoral Fellows**

M.S. McPeck (1994–1995) Professor of Statistics, University of Chicago

P. Marjoram (1995–1998) Research Associate Professor of Biostatistics, USC Keck School of Medicine

J. K. Lee (1995–1997) Associate Professor, Biostatistics & Epidemiology, University of Virginia

S. Schbath (1996) Unité Mathématique, Informatique et Génome (MIG), INRA, Jouy, France

A. Pluzhnikov (1996–1998) Research Associate, Cox lab, University of Chicago

S. Kruglyak (1998–2000) Illumina, Inc.

D. Golinelli (2000–2002) Statistics Group, RAND Corporation, Santa Monica

N. Rosenberg (2001–2004) Assistant Professor, Departments of Human Genetics, Biostatistics, and Ecology & Evolutionary Biology, University of Michigan

P. Calabrese (2001–2004) Research Assistant Professor, Program in Molecular and Computational Biology, USC

P. Nicolas (2004–2005) Unité Mathématique, Informatique et Génome (MIG), INRA, Jouy, France

C. Rangel Escareño (2003–2006) Investigador Asociado, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City

A. E. Teschendorff (Cambridge, 2005–2006) Heller Fellow, UCL Cancer Institute, London

J. Aparicio (2004–2007) CHLA

N. P. Thorne (Cambridge, 2004–2008) Bioinformatics Division, WEHI, Melbourne

I. Tiemann-Boege (Cambridge, 2007–2008) Institute of Biophysics, Johannes Kepler University, Austria

M. E. Ritchie (Cambridge, 2006–2008) Bioinformatics Division, WEHI, Melbourne

N. Barbosa-Morais (Cambridge, 2006–present)

A. G. Lynch (Cambridge, 2006–present)

C. N. Curtis (Cambridge, 2007–present)

S. Samarajiwa (Cambridge, 2008–present)

Q. A. Fu (Cambridge, 2008–present)

B. Carvalho (Cambridge, 2010–present)

### Scientific Officer

I. Spiteri (Senior Scientific Officer; Cambridge, 2009 – present)

### Postgraduate Students

#### Completed PhDs

P. J. Joyce (PhD 1988) *Age-ordered distributions for population genetics models*. Professor, Departments of Mathematics and Statistics, University of Idaho.

R. S. Lundstrom (PhD 1990) *Stochastic models and statistical methods for DNA sequence data*. Idaho Technology Inc., Salt Lake City, Utah.

L. Markovtsova (PhD 2000) *Markov chain Monte Carlo methods in population genetics*. Bioinformatician, Roche Pharmaceuticals.

O. Will (PhD 2001) *Statistical inference in the fossil record*. OTX, Los Angeles, California.

J. Wu (PhD 2001) *Statistical inference for molecular data: man, motifs and microarrays*. Research Associate, Department of Statistics, Carnegie Mellon University.

K. Millman (PhD 2005) *Molecular and genetic epidemiology of Chlamydia trachomatis in the United States*. Postdoc, Dean Lab, CHORI.

D. Abdueva (PhD 2006) *Probe-target hybridization modeling and its application to the analysis of microarrays*. Affymetrix

- D. Scvortsov (PhD 2006) *Detection of deleted and duplicated genomic DNA using HMM analysis of GeneChip data.* Illumina
- W. Xu (PhD 2006) *Searching for and beyond replication origins.* Biostatistician, Stanford Genome Center.
- O. Sargsyan (PhD 2006) *Analytical and simulation results for the general coalescent.* Postdoc, Wakeley Lab, Harvard.
- V. Plagnol (PhD 2006) *Ancestral inference from molecular and paleontological data.* Postdoc, Todd Lab, Cambridge.
- R. Jiang (PhD 2006) *Modeling and analyzing novel genetic data: single feature polymorphisms and polony sequences.* Postdoc, Wall Lab, UCSF.
- R. Alaghband (PhD 2007) *Inference for stochastic models of molecular data.* Trading and Quantitative Research, JPMorgan.
- C. Curtis (PhD 2007) *Analysis of high-density oligonucleotide gene expression data for dissecting aging pathways.* Postdoc, Tavaré lab, CRUK CRI.
- J.-C. Marionni (PhD 2007) *Statistical methods for array CGH and copy number variation experiments.* Postdoc, Stephens Lab, University of Chicago.
- R. Wilkinson (PhD 2007) *Bayesian inference of primate divergence times.* Lecturer, Statistics Department, University of Nottingham.
- M. Dunning (PhD 2008) *Genome-wide analyses using bead-based microarrays.* Bioinformatics Core, CRUK CRI.
- T. Hardcastle (PhD 2009) *Model-driven analysis of high-throughput genomic data in late-stage ovarian cancer.* [Joint with J. Brenton] Postdoc, Baulcombe lab, Cambridge.
- I. Manolopoulou (PhD 2009) *A Bayesian approach to Nested Clade Analysis.* Postdoc, SAMSI, North Carolina.
- D. Grover (PhD 2009) *Real-time tracking and analysis of Drosophila behavior and gene expression.* Postdoc, Rubin lab, HHMI Janelia Farm.
- L. Goldstein (PhD 2009; examined Jan 2010) *Statistical analysis of microRNA expression and related data.*
- C. Spyrou (PhD 2009; examined Dec 2009) *Development and application of Bayesian methodology for some missing data problems in biology.*

## **Current PhDs**

D. Speed (PhD 2006–2009)

S. Knott (PhD 2006–present) [Joint with O. Aparicio]

S. Ivakhno (PhD 2007–2010)

J. Woolford (PhD 2007–2010) [Joint with E. A. Miska]

A. Sottoriva (PhD 2008–2011)

M. L. Smith (PhD 2009–2011)

T. Nellimoottil (PhD 2007–present) [Joint with O. Aparicio]

Supervised 26 MS thesis students.

## **Conference Organization (since 2005)**

Program Committee, RECOMB 2008.

Program Committee, RECOMB 2007.

Chair of Organizing Committee, Program on Stochastic Computation in the Biological Sciences. Isaac Newton Institute, Oct – Dec 2006.

Scientific Committee, 31st conference on Stochastic Processes and their Applications (Bernoulli Society). Paris, France, July 2006.

Organizer, session on Statistical Genetics. 25th European Meeting of Statisticians. Oslo, Norway, July 2005.

Organizing Committee, Workshop on Genome-wide Association Studies. USC, California. April 2005.

## **Invited Lecture Series (since 2005; from 14 since 1990)**

*Copy number variation in the genome: technologies, statistics and applications.* Interface 2007: the 39th Symposium on the Interface of Statistics, Computing Science, and Applications. (4 1-hour lectures, with JC Marioni) Philadelphia, 5/23-26/07.

*Ancestral inference.* 2nd Cornell Probability Summer School on *Probability problems that arise from genetics.* (6 1.25-hour lectures) Cornell, NY. 6/26-7/7/06.

### Invited Conference Lectures (since 2005; from 100 since 1990)

Statistical Methods for Very Large Data Sets. Baltimore, MD. 6/1–3/11.

Invited session on ABC Methods for Genetic Data. 19th International Conference on Computational Statistics (COMPSTAT2010). Paris, France. 08/22–27/10.

*Approximate Bayesian Computation: Theory and biological examples.* Closing Lecture, European Meeting of Statisticians. Piraeus, Greece. 08/17–22/10.

Keynote Lecture, International Conference on Artificial Intelligence and Statistics (AISTATS 2010). Sardinia. 05/13–15/10.

Plenary Lecture, Science at the Shine Dome Annual Symposium: Genomics and Mathematics. Canberra, Australia. 05/07/10.

*Approximate Bayesian Computation and the evolution of tumours.* Conference at Department of Medical Epidemiology and Biostatistics (part of the 200th Anniversary celebration of the Karolinska Institutet). Stockholm, Sweden. 03/26/10.

*Evolutionary approaches to stem cell and tumour dynamics.* Conference on Quantitative Biology and Bioinformatics in Modern Medicine. Queen's University Belfast. 3/2–3/10.

*Evolutionary approaches to cancer.* 26th Genes and Cancer Symposium, University of Warwick. 12/14–16/09.

*Evolutionary approaches to cancer.* Conference on Probability Theory in honour of Professor Andrew Barbour. Zurich, Switzerland. 12/4/09.

*Evolutionary approaches for studying cancer stem cells.* Symposium on Biostatistics and Statistical Genetics. University of Limerick. 9/18/09.

*Some statistical issues in the analysis of Illumina sequencing experiments.* EPSRC Symposium Workshop on Information extraction from complex data sets (INF). Warwick University. 9/14–18/09.

*I spy with my little eye: tracking applications for Drosophila behavioural genetics.* Invited session on Bioinformatics. RSS Conference, Edinburgh. 9/8–11/09.

*Approximate Bayesian Computation: what, why and when?* Invited session on Advanced Monte Carlo Methods. JSM, Washington DC. 8/1–6/09.

*O fly where art thou?* 5th CCBI Annual Meeting. University of Cambridge. 5/20/09.

*O fly where art thou?* Second Biennial Newcastle Workshop on Statistical Bioinformatics and Stochastic Systems Biology. Newcastle University. 5/18–19/09.

*Random trees in biology and cancer.* Tree workshop, Oberwolfach Mathematics Institute. 1/18–24/09.

*Genes and flowering times.* Beyond Boundaries: from Physics to Plant Sciences. A symposium to celebrate 150th anniversary of the birth of JC Bose. Christ's College, Cambridge. 12/6/08.

*O stem cell, where art thou?* Minisymposium on Mathematical Analysis of Biological Sequences, SIAM 2008. San Diego, CA. 7/7–11/08.

*High-throughput analysis of methylation patterns for tracking cell lineages.* Cambridge Cancer Centre Symposium, Cambridge UK. 6/27/08.

*Computational molecular biology: an introduction for statisticians.* RSS Medical Section Meeting “Introduction to Bioinformatics”. Royal Statistical Society. London. 6/26/08.

*High-throughput detection of methylation patterns for tracking cell lineages.* Keynote address. Emerging Statistical Challenges in Genome and Translational Research. Banff International Research Station, Banff, Canada. 6/1-6/08.

*Using methylation tags to track stem cells.* Stem Cells 2007. Punta Cana, Dominican Republic. 12/13-16/07.

*Statistical issues in copy number detection.* BioC2007. Seattle, WA. 8/6-7/07.

*Stochastic models for the evolution of stem cells in colon crypts.* 26th Leeds Annual Statistical Research (LASR) Workshop. Leeds, UK. 7/4-6/07.

*Statistical issues when detecting copy number variation.* Statistics for Biomolecular Data Integration and Modeling. Ascona, Switzerland. 6/10-15/07.

*O stem cell, where art thou?* CBMI Annual Symposium. University of Kent. 4/26/07.

*Inferring the behaviour of colon crypts by exploiting methylation patterns.* Workshop on Statistics in Genomics and Proteomics. Coimbra, Portugal. 3/9-10/07.

*Oligo antics: array applications in cancer research.* COGEME Interest Group Meeting. Manchester, UK. 11/7-8/06.

*Now you know your ABCs: examples and problems.* Workshop on Recent Advances in Monte Carlo based inference. Isaac Newton Institute, Cambridge UK. 10/30/06.

*Some comments on approximate Bayesian computation* XXIII International Biometric Conference. Montreal. 7/16-21/06.

*Using methylation patterns to understand the stem cell population of the human colon.* Bioinformatics 2006. Aarhus, Denmark. 6/15-18/06.

*Whole genome scans: full employment for statisticians?* East Kent Section, Royal Statistical Society. Canterbury, Kent. 4/05/06.

*Genome-wide associations of gene expression variation in humans.* Sequence Analysis Toward System Biology Workshop. IPAM, Los Angeles. 1/9-13/06.

*Statistical issues for expression analysis of Illumina bead-based microarrays.* Plenary Lecture, Workshop on Statistics in Genomics and Proteomics. Estoril, Portugal. 6-8/10/05.

*Posterior inference on the stem cell population of the human colon through analysis of methylation patterns.* Session on Bioinformatics, JSM2005. Minneapolis, MN. 7-11/8/05.

*Using the fossil record to date splits in the primate tree.* Stochastic Modelling of Complex Systems (SMOCS-05). Great Barrier Reef, Australia. 10-16/7/05.

*Identifying expression variation with the Illumina platform.* Keynote Speaker, International Workshop on Statistical Modelling (focus on statistical contributions to modern genomic research). Sydney, Australia. 10-15/7/05.

*Some comments on approximate Bayesian computation.* Keynote Speaker, Australian Mathematical Sciences Institute Symposium on Recent Advances in Biostatistics, Bioinformatics and Markov chain Monte Carlo. Sydney, Australia. 8-9/7/05.

*Stem cell evolution in the colon.* Symposium on Branching in Biology. Gothenburg University, Sweden. 5/22-25/05

*Identifying expression variation on the Illumina platform.* International HapMap Project conference on Genomic Studies and the HapMap. University of Oxford, UK. 3/15-18/05

### **Contributed Conference Lectures or Departmental Seminars**

33 (since 2005; from 157 since 1990) in Australia, Canada, Denmark, France, Germany, Ireland, Sweden, Switzerland, UK, USA.

January 29, 2010