

CURRICULUM VITÆ

Sophie SCHBATH

INRA, French National Institute of Agronomic Research

BIOGRAPHICAL

Born December 19, 1969 in Nantes, France.

Married with 2 children.

Citizenship: France.

E-mail: Sophie.Schbath@jouy.inra.fr

EDUCATION

- 2003 Habilitation, University of Evry, France. *Two mathematical approaches to genome analysis: statistic of word counts and prediction in physical mapping.*
- 1995 Ph.D. in Statistics, University of Paris V, France. *Étude asymptotique du nombre d'occurrences d'un mot dans une chaîne de Markov et application à la recherche de mots de fréquence exceptionnelle dans les séquences d'ADN.*
- 1992 Master degree in Stochastic modeling and Statistics, University of Paris XI.

PROFESSIONAL EXPERIENCE

- 2006 – Directrice de Recherche, Unit of Mathématique, Informatique & Génome, INRA, Jouy-en-Josas.
- 2000 – 2005 Chargée de Recherche, Unit of Mathématique, Informatique & Génome, INRA, Jouy-en-Josas.
- 1996 – 1999 Chargée de Recherche, Unit of Biométrie, INRA, Jouy-en-Josas.
- 1996 Post-doc, Mathematics Department, University of Southern California, Los Angeles, USA.
- 1992 – 1996 Attachée Scientifique Contractuelle, Unit of Biométrie, INRA, Jouy-en-Josas.

REFEREEING SERVICES and SCIENTIFIC COMMITTEES

Referee for *Annals of Applied Probability* (2 papers), *Journal of Applied Probability* (4 papers), *ESAIM: Probability and Statistics* (1 paper), *Combinatorics, Probability and Computing* (1 paper), *Methodology and Computing in Applied Probability* (2 papers), *Discrete Applied Mathematics* (2 papers), *Annales de l'Institut Henri Poincaré* (1 paper), *Annals of the Institute of Statistical Mathematics* (1 paper), *Biometrics* (1 paper), *Statistical Applications in Genetics and Molecular Biology* (4 papers), *Journal of Mathematical Biology* (1 paper), *Journal*

of *Computational Biology* (7 papers), *Bioinformatics* (1 paper), *BMC Bioinformatics* (1 paper), *IEEE Transactions on Computational Biology and Bioinformatics* (2 papers), *Journal of Computer & Chemistry* (1 paper), *INFORMS-Journal of Computing* (1 paper), *Nucleic Acid Research* (1 paper), *Genomics* (2 papers), *Journal of Molecular Evolution* (1 paper), *Canadian Journal of Microbiology* (1 paper), *Archae* (1 paper).

Member of Program Committees of conferences:

- *3rd Annual International Conference on Computational Molecular Biology* (RECOMB), Lyon, France. April 1999.
- *2nd Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Toulouse, France. May 2001.
- *3rd Workshop on Algorithms in Bioinformatics* (WABI), Budapest, Hungary. September 2003.
- *8th Annual International Conference on Computational Molecular Biology* (RECOMB), San Diego, California, USA. March 2004.
- *6th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Lyon, France. July 2005.
- *5th Workshop on Algorithms in Bioinformatics* (WABI), Eivissa, Spain. October 2005.
- *7th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Bordeaux, France. July 2006.
- *11th Annual International Conference on Computational Molecular Biology* (RECOMB), Oakland, California, USA. April 2007.

Member of Scientific Committees:

- French *Bioinformatics Program* (2000 – 2003).
- French ACI *IMPBio* (2003 - 2004).
- INRA Department *Mathématiques et Informatique Appliquées* (2002 – 2010).
- INRA Evaluation section *Mathématique, Bioinformatique et Intelligence Artificielle* (2002 – 2010).

Member of Ph.D. committees:

- Élodie Nédélec, University Paris XI, member (2004).
- Gaëlle Gusto, University Paris XI, advisor (2004).
- Christelle Melo de Lima, University of Lyon, reviewer (2005).
- Narjiss Touyar, university of Rouen, co-advisor (2006).
- Leonor Palmeira, University of Lyon, reviewer (2007).
- Etienne Roquain, University Paris XI, advisor (2007).
- Fabrice Touzain, University of Nancy, member (2007).
- Fanny Villers, University Paris XI, member (2007).

CONFERENCE ORGANIZATION _____

Co-organizer of the 11th *European Young Statisticians Meeting*, Marly-le-Roi, France (<http://www.inra.fr/bia/J/SCS/EYSM/welcome.html>). August 1999.

Local Organizing Committee, *3rd Annual International Conference on Computational Molecular Biology* (RECOMB), Lyon, France. April 1999.

Co-organizer of the workshop “Mathematics for Biological Networks”, Paris, France (<http://stat.genopole.cnrs.fr/MBN2007/>). December 2007.

Organization of sessions:

“Phylogénie” and “Analyse statistique des séquences” in the *Journées MAS : Modélisation pour les Sciences du Vivant*, Grenoble, France. September 2002.

- “Probabilistic problems on words in computational biology” in the *International Workshop in Applied Probability*, University of Piraeus, Greece. March 2004.

- “Probability and Statistics applied to Computational Biology” in the *International Workshop in Applied Probability*, University of Connecticut, USA. May 2006.

- “Applied probability methodology in computational biology” in the 31st *Conference on Stochastic Processes and their Applications* Paris, France. July 2006.

PUBLICATIONS

Journal Articles

- [1] SCHBATH, S. (1995). Compound Poisson approximation of word counts in DNA sequences. *ESAIM: Probability and Statistics*. **1** 1–16.
- [2] SCHBATH, S., PRUM, B. and TURCKHEIM, É. DE (1995). Exceptional motifs in different Markov chain models for a statistical analysis of DNA sequences. *J. Comp. Biol.* **2** 417–437.
- [3] SCHBATH, S. (1997). Coverage processes in physical mapping by anchoring random clones. *J. Comp. Biol.* **4** 61–82.
- [4] SCHBATH, S. (1997). An efficient statistic to detect over- and under-represented words in DNA sequences. *J. Comp. Biol.* **4** 189–192.
- [5] REINERT, G. and SCHBATH, S. (1998). Compound Poisson and Poisson process approximations for occurrences of multiple words in Markov chains. *J. Comp. Biol.* **5** 223–254.
- [6] EL KAROUI, M., BIAUDET, V., SCHBATH, S. and GRUSS, A. (1999). Characteristics of Chi distribution on several bacterial genomes. *Research in Microbiology*. **150** 579–587.
- [7] REINERT, G., SCHBATH, S. and WATERMAN, M. (2000). Probabilistic and statistical properties of words: an Overview. *J. Comp. Biol.* **7** 1–46.
- [8] SCHBATH, S., BOSSARD, N. and TAVARÉ, S. (2000). The effect of non-homogeneous clone length distribution on the progress of an STS mapping project. *J. Comp. Biol.* **7** 47–58.
- [9] SCHBATH, S. (2000). An overview on the distribution of word counts in Markov chains. *J. Comp. Biol.* **7** 193–201.
- [10] ROBIN, S. and SCHBATH, S. (2001). Numerical comparison of several approximations of the word count distribution in random sequences. *J. Comp. Biol.* **8** 349–359.

- [11] ROBIN, S., DAUDIN, J.-J., RICHARD, H., SAGOT, M.-F. and SCHBATH, S. (2002). Occurrence probability of structured motifs in random sequences. *J. Comp. Biol.* **9** 761–773.
- [12] SCHBATH, S. (2003). Statistical methods in physical mapping. *Encyclopedia of the Human Genome*, 434, Nature Publishing Group. (<http://www.ehgonline.net/mathematical.asp>)
- [13] SCHBATH, S. (2004). A la recherche de mots de fréquence exceptionnelle dans les génomes. *Images des Mathématiques*.
- [14] GUSTO, G. and SCHBATH, S. (2005). FADO: a statistical method to detect favored or avoided distances between motif occurrences using the hawkes’ model. *Statistical Applications in Genetics and Molecular Biology*. **4**, Article 24.
- [15] MATIAS, C., SCHBATH, S., BIRMELÉ, E., DAUDIN, J.-J. and ROBIN, S. (2006). Network motifs: mean and variance for the count. *REVSTAT*. **4** 31–51.
- [16] STEFANOV, V., ROBIN, S., and SCHBATH, S. (2007). Waiting times for clumps of patterns and for structured motifs in random sequences. *Discrete Applied Mathematics*. **155**, 868–880.
- [17] ROQUAIN, E. and SCHBATH, S. (2007). Improved compound Poisson approximation for the number of occurrences of multiple words in a stationary Markov chain. *Adv. Appl. Prob.* **39** 1–13.
- [18] ROBIN, S., SCHBATH, S. and VANDEWALLE, V. (2007). Statistical tests to compare motif count exceptionalities. *BMC Bioinformatics* **8:84**, 20 pages.
- [19] HALPERN, D., CHIAPELLO, H., SCHBATH, S., ROBIN, S., HENNEQUET-ANTIER, C., GRUSS, A. and EL KAROUI, M. (2007). Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modelling. *PLoS Genetics*. **3(9)** e153.
- [20] TOUZAIN, F., SCHBATH, S., DEBLED-RENNESON, I., AIGLE, B., LEBLOND, P. and KUCHEROV, G. (2008). SIGffRid: a tool to search for σ factor binding sites in bacterial genomes using comparative approach and biologically driven statistics. *BMC Bioinformatics*. **9:73** 1–23.
- [21] PICARD, F., DAUDIN, J.-J., KOSKAS, M., SCHBATH, S. and ROBIN, S. (2008). Assessing the exceptionality of network motifs. *J. Comp. Biol.* **15:1** 1–20.
- [22] TOUYAR, N., SCHBATH, S., CELLIER, D. and DAUCHEL, H. (2008). Poisson approximation for the number of repeats in a Markov chain model. To appear in *J. Appl. Prob.*

Books / Book Chapters

- [23] SCHBATH, S. and BOUVIER, A. (1998). Finding words with unexpected frequencies in DNA sequences. In *Explorapedia of Statistical and Mathematical Techniques for use in Research and Technology* (<http://www.bioss.sari.ac.uk/smart/unix/intro/slides/home.htm>).

- [24] REINERT, G. and SCHBATH, S. (1999). Compound Poisson approximations for occurrences of multiple words. In *Statistics in Genetics and Molecular Biology*, (F. Seillier, ed.). IMS Lecture Notes-Monograph Series. **33** 257–275.
- [25] ROBIN, S., RODOLPHE, F. and SCHBATH, S. (2003). *ADN, mots et modèles*. BELIN.
- [26] REINERT, G., SCHBATH, S. and WATERMAN, M. (2005). Statistics on words with applications to biological sequences. In *Applied Combinatorics on Words*, (J. Berstel and D. Perrin, ed.). Cambridge University Press.
- [27] ROBIN, S., RODOLPHE, F. and SCHBATH, S. (2005). *DNA, Words and Models*. Cambridge University Press.

Technical Reports

- [28] GÉLIS, F. and SCHBATH, S. (1996). *R'MES : Recherche de Mots Exceptionnels dans les Séquences d'ADN – Version 1*. Notice d'utilisation. INRA, Biométrie, 78352 Jouy-en-Josas, France.
- [29] BOUVIER, A., GÉLIS, F. and SCHBATH, S. (1999). *R'MES : Recherche de Mots Exceptionnels dans les Séquences d'ADN – Version 2*. Guide de l'utilisateur. INRA, Biométrie, F78352 Jouy-en-Josas.
- [30] EL KAROUI, M. and SCHBATH, S. (2001). Identification de motifs significativement sur- ou sous-représentés dans un génome : le cas de **gctggtgg** dans le génome d'*Escherichia coli*. Rapport technique pour illustrer un cours de Probabilités de l'Ecole Polytechnique (10 pages).
- [31] SCHBATH, S. (2006). Statistics of motifs. Lecture notes for *Atelier INSERM Identification of non-coding functional regions in genome*, La Londe-les-Maures, April 27-28 (10 pages). <http://migale.jouy.inra.fr/outils/mig/rmes/atelier-inserm2006.pdf>
- [32] HOEBEKE, M. and SCHBATH, S. (2006). *R'MES: Finding exceptional motifs, version 3*. User guide. <http://migale.jouy.inra.fr/outils/mig/rmes/rmes3.01.userGuide.pdf>

Invited Conference Lectures

- [33] SCHBATH, S. (1997). Predicting progress in a physical mapping project by anchoring random clones using coverage processes. Invited Lecture in *Mathematical Statistics and its Application to Biosciences*. ISI Satellite Meeting, Rostock, Germany. Aug 31 - Sept 4.
- [34] SCHBATH, S. (2000b). (i) Introduction to the problem of finding words with unexpected frequency in DNA sequences: motivation, word counts, models, periodic structure of words, word count distribution, (ii) Gaussian approximation of the word count distribution and application, (iii) Poisson approximation and the chen-stein method. In *Semester in Bioinformatics*. Uppsala, Sweden. February 24-25.

- [35] SCHBATH, S. (2000). Modèles markoviens dans l'analyse statistique des séquences. Invited Lecture in *Journées TAS, Traitement et Analyse de Séquences*. Evry, France. Nov 22-24.
- [36] SCHBATH, S. (2001). Distribution of word counts in DNA sequences and quality of approximations. Invited Lecture in *23rd European Meeting of Statisticians*. Funchal, Madeira. Aug 13-18.
- [37] SCHBATH, S. (2002). Exceptional motifs in biological sequences. In *Maps, Sequences and Genomes*. University of Southern California, Los Angeles, USA. May, 31 – June, 2 (abstract).
- [38] SCHBATH, S. (2004). Overview on probabilistic problems on words in computational biology. In *International Workshop in Applied Probability*. University of Piraeus, Greece. March, 22-25.
- [39] SCHBATH, S. (2004). Modèles statistiques et analyse de génomes. In *36èmes Journées de Statistique*. Montpellier, France. May, 22-28.
- [40] SCHBATH, S. (2004). Modelling the dependence between sequence motifs. In *6th World Bernoulli Congress*. Barcelona, Spain. July, 26-30.
- [41] SCHBATH, S. (2005e). Statistical problems arising in physical mapping. In *Workshop on Stat. in Genomics and Proteomics*. Estoril, Portugal. October, 6-8.
- [42] SCHBATH, S. (2005f). The statistical world of motifs on genomes. In *SemStat Summer School*. University of Warwick, UK. September, 11-13.
- [43] SCHBATH, S. (2006b). Statistics of motifs. In *Atelier INSERM Identification de régions non codantes fonctionnelles dans les génomes*. La Londe-les-Maures. April, 27-28.
- [44] SCHBATH, S. (2006a). Network motifs: mean and variance for the count. In *International Workshop on Applied Probability*. University of Connecticut, USA. May, 15-18.
- [45] SCHBATH, S. (2007). Assessing the exceptionality of network motifs. In *Workshop on Statistics in Genomics and Proteomics*. Centro Internacional de Matemática, Coimbra, Portugal. March, 9-10.

Conference Lectures

- [46] SCHBATH, S. (1994). Identification de motifs exceptionnels par l'étude statistique des comptages de "trains". In *Recherche de Motifs dans les Séquences*. GREG et GDR "Informatique et Génomes", Marseille, France. February 24-25 (abstract).
- [47] SCHBATH, S. (1994). Étude des comptages de mots dans des séquences d'ADN et approximations par des lois de Poisson. In *XXVIèmes Journées de Statistique*. Association pour la Statistique et ses Utilisations, Neuchâtel, Switzerland. May 24-27 (abstract).
- [48] SCHBATH, S. (1994). Recherche de motifs de fréquence exceptionnelle dans les séquences d'ADN. In *Forum InterDisciplinaire "Génome et Informatique"*. GREG et GDR "Informatique et Génomes", Aussois, France. June 15-17 (abstract).

- [49] SCHBATH, S. (1995). Statistiques des comptages de mots dans les séquences d'ADN. In *XXVIIèmes Journées de Statistique*. Association pour la Statistique et ses Utilisations, Jouy-en-Josas, France. May 15-19 (abstract).
- [50] SCHBATH, S. (1995). Statistics of counts of words in DNA sequences. In *21st European Meeting of Statisticians*. Aarhus University, Denmark. August 21-25 (abstract).
- [51] SCHBATH, S. (1997). Predicting progress in physical mapping projects without homogeneity assumptions. In *Statistics and Inference in Molecular Biology*. Program in Mathematics and Molecular Biology, Santa Fe, USA. January 14-19 (abstract).
- [52] SCHBATH, S. (1998). Approximation for counts of multiple words in biological sequences. In *Workshop on Mathematical and Statistical Aspects of Molecular Biology*. University of Wales College of Medicine, Cardiff, GB. April 6-7 (abstract).
- [53] SCHBATH, S. (1998). How to find exceptional words in biological sequences. In *Analyse Structurale et Fonctionnelle d'un génome*. Séminaire Algorithmes et Biologie, Institut Pasteur, Paris, France. December 8-10.
- [54] SCHBATH, S. (1999). A new method for protein classification based on motifs. In *Electrophoresis Forum'99*. Rouen, France. November 24-26 (abstract).
- [55] SCHBATH, S. (2000). Finding protein interactions by analyzing occurrences of multiple motifs along DNA sequences. In *Third Danish-French workshop on spatial statistics and image analysis in biology*. Luminy, France. March 7-10.
- [56] SCHBATH, S. and GUSTO, G. (2002). Analyse statistique de la corépartition de motifs le long d'une séquence. In *Journées MAS : Modélisation pour les Sciences du Vivant*. Grenoble, France. September, 2-4 (abstract).
- [57] SCHBATH, S. (2002b). Statistiques des comptages de mots dans les séquences. In *Séminaire Algorithmes et Biologie*. Lyon, France. October, 1-3 (abstract).
- [58] BOURGAIN, I., CHIAPELLO, H., HENNEQUET-ANTIER, C., ROBIN, S., SCHBATH, S., GRUSS, A. and EL KAROUI, M. (2003). Genomic distribution of short motifs involved in DNA repair in pathogenic and non pathogenic *E. coli* In *Second European Conference on Computational Biology*. Paris, France. September, 27-30 (**selected short paper**, 7-9).
- [59] SCHBATH, S. (2005b). La recherche de mots de fréquence exceptionnelle dans une séquence : état de l'art et derniers résultats. In *Séminaire Probabilités, Optimisation, Contrôle*. INRIA Rocquencourt, France. January, 20 (abstract).
- [60] SCHBATH, S. (2005g). À la recherche de motifs exceptionnels dans les génomes. In *Séminaire Mathématiques Appliquées*. Université Paris V, France. April, 15 (abstract).
- [61] SCHBATH, S. (2005h). À la recherche de motifs exceptionnels dans les génomes : exemples de problèmes statistiques. In *Séminaire Modèles Stochastiques*. Ecole Polytechnique, Palaiseau, France. May, 23 (abstract).
- [62] TOUZAIN, F., SCHBATH, S., DEBLED-RENNESON, I., AIGLE, B., LEBLOND. and KUCHEROV, G. (2005). SIGffRid : Programme de recherche des sites de fixation des facteurs de transcription par approche comparative. In *Journées Ouvertes Biologie Informatique Mathématiques*. Lyon, France. July, 6-8 (**selected long paper**, 417-426).

- [63] SCHBATH, S. (2005c). Modeling the dependence between sequence motifs. In *Journées Algorithmique Génomique*. University of Paris XI, Orsay, France. November, 24-25 (abstract).
- [64] SCHBATH, S. (2005a). Approximation de poisson pour le nombre de répétitions dans une séquence markovienne. In *Journées Algorithmique Génomique*. University of Paris XI, Orsay, France. November, 24-25 (abstract).
- [65] SCHBATH, S. (2005d). Network motifs: mean and variance for the count. In *Comp-BioNet*. Lyon, France. December, 5-7 (paper).
- [66] SCHBATH, S. (2006c). Towards exceptional motifs in biological networks. In *Workshop Statistical Methods for Post-Genomics data*. INSA, Toulouse, France. March, 30-31 (abstract).
- [67] PICARD, F., DAUDIN, J.-J., SCHBATH, S. and ROBIN, S. (2006a). Assessing the exceptionality of network motifs. In *Journée thématique Réseaux d'interactions : analyse, modélisation et simulation*. Université de Lyon, France. Octobre, 29 (short paper).
- [68] SCHBATH, S. (2006a). Comment tester qu'un motif est significativement plus exceptionnel dans une séquence que dans une autre ? In *Groupe de Travail en Génomique Comparative*. Université de Nantes, France. October, 12-13 (abstract).
- [69] SCHBATH, S. (2007b). Assessing the exceptionality of network motifs. In *Tripartite meeting BIOS/INRA/Biometris*. University of Wageningen, The Netherlands (abstract).
- [70] PICARD, F., DAUDIN, J.-J., KOSKAS, M., SCHBATH, S. and ROBIN, S. (2007). Assessing the exceptionalities of network motifs. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (C. Brun and G. Didier, ed.), Marseille, France (**selected long paper**, 235–241).
- [71] ROBIN, S., SCHBATH, S. and VANDEWALLE, V. (2007). Statistical tests to compare motif count exceptionalities. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (C. Brun and G. Didier, ed.), Marseille, France (**selected long paper**, 57–62).

STUDENTS

Postgraduate Students

GUSTO, G. (Dec 2004). *Estimation dans un modèle de Hawkes et application à l'étude de la corépartition de motifs le long d'un génome*, University Paris XI. Co-advisor: P. Massart.

TOUYAR, N. (Jan 2006). *Approximation de la loi du nombre de répétitions dans une chaîne de Markov*, University of Rouen. Co-advisors: D. Cellier and H. Dauchel.

ROQUAIN, E. (Oct 2007). *Motifs exceptionnels dans des séquences hétérogènes. Contributions à la théorie et à la méthodologie des tests multiples*, University Paris XI.

DEVILLERS, H. (started November 2007). *Statistics of complete genome alignments*, University of Evry.

Graduate Students

- BOSSARD, N. (1997). *Prédiction en cartographie physique par la méthode d'ancrage, en l'absence d'homogénéité sur la longueur des clones*. Master's thesis, D.E.S.S. Informatique Appliquée à la Biologie, Université Versailles-Saint-Quentin (6 months).
- LEPAGE, F. (1999). *Implémentation et analyse d'une nouvelle méthode statistique de classification de protéines*. Master's thesis, D.E.S.S. Informatique Appliquée à la Biologie, Université Versailles-Saint-Quentin (6 months).
- NÉDÉLEC, É. (1999). *Recherche des mots exceptionnels dans les séquences d'ADN conditionnellement à l'arrivée d'un long mot*. Master's thesis, Modélisation Stochastique et Statistique, Université Paris-XI, Orsay (4 months).
- GUSTO, G. (2000). *Approximation par une loi de poisson composée de la loi du comptage d'un mot rare dans une chaîne de markov*. Master's thesis, D.E.A. Modélisation Stochastique et Statistique, Université Paris-XI, Orsay (3 months).
- BASTIÈRE, J. (2002). *Etude de la répartition des sites Chi sur le génome de plusieurs souches de Escherichia coli*. Master's thesis, Maîtrise de Biologie Cellulaire et Physiologie, Université de Versailles. (1 month).
- LAJUS, A. (2002). *Conception d'une interface JAVA pour la comparaison de l'exceptionnalité des mots dans deux séquences d'ADN*. Master's thesis, 3ème année IUP Génie Biologique et Informatique, Université d'Evry. (7 months).
- WYNANT, W. (2002). *Modélisation de la corépartition de mots le long d'un génome*. Master's thesis, DESS Ingénierie Mathématique - Université Paris XI. (6 months).
- DEMIZIEUX, R. (2002). *Etude de la significativité statistique des scores obtenus à partir de la méthode de reconnaissance de repliement des protéines FROST*. Master's thesis, DEA Statistique et modèles aléatoires en Economie et Finances, Université Paris VII. (4 months).
- WYNANT, W. (2003). *Recherche de mots exceptionnels dans une chaîne de Markov cachée*. Master's thesis, DEA Application des Mathématiques et de l'Informatique à la Biologie - Université d'Evry (6 months).
- BOURGAIT, I. (2003). *Comparaison d'un système de réparation de l'ADN chez des souches pathogènes et non pathogènes d'E. coli*. Master's thesis, DESS Ingénierie et Génomique Fonctionnelle - Université de Paris 7 (6 months).
- SIMON, G. (2003). *Analyse statistique du contexte des motifs Chi chez E. coli*. Master's thesis, 1ère année de l'ENSAE - Malakoff (5 weeks).
- GUÉRIN, J. (2004). *Réalisation d'une interface java pour les résultats du logiciel R'MES*. Master's thesis, DESS Informatique Appliquée à la Biologie - Universités Paris VI, Versailles et Evry. (5 months).
- ROQUAIN, E. (2004). *Approximation de Poisson composée du comptage d'une famille de mots dans une chaîne de Markov*. Master's thesis, DEA Modélisation Stochastique et Statistique - Université d'Orsay. (5 months).

GROSZ, S. (2006). *Indice de confiance pour les boucles spécifiques d'un génome obtenues par alignement de génomes complets*. Master's thesis, BioInformatique et BioStatistiques - Université d'Orsay. (6 months).

MISCELLANEOUS

Co-director (with A. Denise and C. Thermes) of the French GdR CNRS 3003 *BioInformatique Moléculaire* which gathers the French community in bioinformatics [around a thousand people] (2006–2009). <http://www.gdr-bim.u-psud.fr/>

Leader of the Research Group *Statistics for Systems Biology* (1995 –).
See <http://genome.jouy.inra.fr/ssb/>

Co-author (with Gélis, F., Bouvier, A., Hoebeke, M.) of the *R'MES* software dedicated to the detection of exceptional words in biological sequences (freely available at <http://genome.jouy.inra.fr/ssb/rmes/>)

Teaching activity on statistical methods in genome analysis (41 hours in 2000, 46 hours in 2001, 49 hours in 2002, 47 hours in 2003, 44 hours in 2004, 28 hours in 2005, 48 hours in 2006, 24 hours in 2007).

November 7, 2007.