CURRICULUM VITÆ

Sophie SCHBATH
INRA, French National Institute for Agricultural Research

BIOGRAPHICAL

Born December 19, 1969 in Nantes, France.
Married with 2 children.
Citizenship: France.
E-mail: Sophie.Schbath@inrae.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

2018 –   Directrice de Recherche 1ère classe, Unit Applied Mathematics and Computer Science, from Genomes to the Environnement (MaIAGE), INRAE, Jouy-en-Josas.

2006 – 2018   Directrice de Recherche 2e classe, Unit of Mathématique, Informatique & Génome (MIG) then MaIAGE INRA, Jouy-en-Josas.

2000 – 2005   Chargée de Recherche, Unit of Mathématique, Informatique & Génome (MIG), 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.

ADMINISTRATION, RESPONSABILITIES

Head of the Inra laboratory *Applied Mathematics and Computer Science, from Genome to the Environnement* (2015 – 2022).

Head of the Inra laboratory *Mathématique, Informatique et Génome* (2012 – 2014).

Scientific head of the bioinformatics platform Migale (2016 –).

Co-director of the French GdR CNRS 3003 *BioInformatique Moléculaire* which gathers the French community in bioinformatics [around a thousand people] (2006–2009, 2010–2013). http://www.gdr-bim.u-psud.fr/

President of the French BioInformatics Society (2010 – 2016).

Leader of the Research Group *Statistics for Systems Biology* (1995 – 2012).
See http://www.ssbgroup.fr

REFEREEING SERVICES and SCIENTIFIC COMMITTEES

Reviewer of 56 journal articles from 1996 for
- *Annals of Applied Probability, Journal of Applied Probability, ESAIM: Probability and Statistics, Combinatorics, Probability and Computing, Methodology and Computing in Applied Probability, Discrete Applied Mathematics, Glasnik Matematicki, Annales de l'Institut Henri Poincarré, Annals of the Institute of Statistical Mathematics, Biometrics, Statistical Applications in Genetics and Molecular Biology,*
- *Journal of Computational Biology, Journal of Mathematical Biology, Bioinformatics, BMC Bioinformatics, IEEE Transactions on Computational Biology and Bioinformatics, Journal of Bioinformatics and Computational Biology, Journal of Computer & Chemistry, INFORMS-Journal of Computing, Europhysics Letters,*
- *Nucleic Acid Research, Genomics, Journal of Molecular Evoluvion, Canadian Journal of Microbiology, Archae.*

Member of Editorial Boards:
- *Scandinavian Journal of Statistics* (Sept 2012 - Aug 2019)
- *Journal of Computational Biology* (2014 - 2019)
- *Methodology and Computing in Applied Probability* (2017 - 2021)
- *Nucleic Acid Research Genomics and Bioinformatics* (April 2019 -)

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, Californie, USA. April 2007.
- *10th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Nantes, France.

June 2009.
- *42th Journées de Statistiques* (JDS), Marseille, France. May 2010.
- *12th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Paris, France. June 2011.
- *13th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Rennes, France. July 2012.
- *14th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Toulouse, France. July 2013.
- *16th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Clermont-Ferrand, France. July 2015.
- *8th International Conference on Bioinformatics Models, Methods and Algorithms* (BIOIN-FORMATICS 2017), Porto, Portugal. February 2017.
- *19th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Marseille, France. July 2018.
- *20th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Nantes, France. July 2019.
- *21th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Montpellier, France. July 2020.
- *22th Journées Ouvertes : Biologie, Informatique, Mathématique* (JOBIM), Paris, France. July 2021. Scientific Chair.

Member of Scientific Committees:
- French *Bioinformatics Program* (2000 – 2003).
- French ACI *IMPBio* (2003 - 2004).
- INRA Division *Mathématiques et Informatique Appliquées* (2002 – 2011).
- INRA *Bioinformatics committee* (2009 – 2016).
- INRA Metaprogramme *Metaomics for Microbial Ecosystems* de l'INRA (2010 – 2018).
- Science Advisory Board of the european project RADIANT (2013 – 2015)
- Science Advisory Board of the Computational Biology Institute at Montpellier (2014 - 2017)
- French GdR CNRS 3003 *BioInformatique Moléculaire* (2014 – 2020)
- Academic Council of University Paris-Saclay (2015 – 2019)
- Mathematics Department of University Paris-Saclay (2016 – 2020)
- Mathematics Graduate School of University Paris-Saclay (2020 –)
- Steering committee of Institut Convergence *INCEPTION* (2017 –)
- Steering committee of Institut Convergence *DataIA* (2019 –)

Member of Evaluation Committee :
- INRA Scientific Evaluation section *Mathématique, Bioinformatique et Intelligence Artificielle* (2002 – 2010).
- AERES evaluation committee of the TIMC laboratory in Grenoble (2010).
- Evaluation committee of Institut Pasteur (2016 – 2019).
- INRAE Scientific Evaluation *Soutien et Pilotage de la Recherche* (2021 – 2024).
- HCERES evaluation committee of the LITIS laboratory in Rouen (2021).

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, president (2007).
- Fanny Villers, University Paris XI, president (2007).
- Aude Liefooghe, University of Lille, member (2008).
- Simona Grusea, University of Marseille, reviewer (2008).
- Elisabeth Ford, Oxford University, reviewer (2009).
- Lisbeth Carstensen, Copenhagen University, reviewer (2010).
- Hugo Devillers, University of Evry, advisor (2011).
- Jean-Baka Domelevo-Entfellner, University of Montpellier, reviewer (2011).
- Abdelkader Behdenna, University Paris Pierre et Marie Curie, member (2016).
- Ibrahim Sultan, University Paris-Saclay, co-advisor (2019).
- Christophe Menichelli, University of Montpellier, member (2019).
- Guillaume Gautreau, University of Evry, member (2020).
- Romain Ménégaux, PSL Research University, Paris, member (2021).

Member of Habilitation committees:
- Valentina Boeva, University Paris 6, member (2014).
- Pierre Peterlongo, University of Rennes, reviewer (2016).
- Guillem Rigaill, University Paris-Saclay, reviewer (2020).
- Rayan Chikhi, Sorbonne University, reviewer (2021).

TEACHING, MEDIATION, OUTREACH

Teaching activity on statistical methods in genome analysis (particularly Markov models and statistics of motifs) or on the R language: about 45 hours per year on 2000-2011, about 24h/year from 2012.

Co-author (with Gélis, F., Bouvier, A., Hoebeke, M.) of the *R'MES* software dedicated to the detection of exceptionnal words in biological sequences (freely available at `http://migale.jouy.inra.fr/outils/mig/rmes`)

Participation in the creation and animation of a workshop for the general public on the sequence alignment problem during the open days of the Inra centre in Jouy-en-Josas on the occasion of the 70th anniversary of INRA (2016).

Interview by the Young French Bioinformaticians (2016).

CONFERENCE ORGANIZATION

Co-organizer of:
- 11th *European Young Statisticians Meeting*, Marly-le-Roi, France. August 1999.
- 3rd *Annual International Conference on Computational Molecular Biology* (RECOMB), Lyon, France. April 1999.
- workshop *Mathematics for Biological Networks*, Paris, France (`http://stat.genopole.cnrs.fr/MBN2007/`). December 2007.
- Journées du *GdR Bioinformatique Moléculaire*, Paris, France (`http://www.gdr-bim.u-psud.`

`fr/journees-gdr.php`). November 2009.

- 1st *Bioinformatics for Regulatory Genomics* ISMB Special Interest Group meeting (BioReg SIG), Boston, USA (`http://light.ece.ohio.edu/bioreg/2010/`). July 2010.

- 2nd *Bioinformatics for Regulatory Genomics* ISMB Special Interest Group meeting (BioReg SIG), Vienna, Austria (`http://light.ece.ohio.edu/bioreg/2011/`). July 2011.

- Journées du *GdR Bioinformatique Moléculaire*, Paris, France (`http://mig.jouy.inra.fr/?q=fr/journees-gdrbim-2012`). January 2012.

- Journées du *GdR Bioinformatique Moléculaire*, Paris, France (`http://www.gdr-bim.cnrs.fr/coll-2013/`). November 2013.

- ECCB'14 satellite workshop *Recent Computational Advances in Metagenomics*, Strasbourg, France. September 2014.

- 13th *European Conference on Computational Biology* (ECCB), Strasbourg, France (`http://www.eccb14.org`). September 2014.

- 2e workshop *Recent Computational Advances in Metagenomics* (RCAM), Paris, France. September 2015.

- 3e workshop *Recent Computational Advances in Metagenomics* (RCAM), as a satellite of ECCB'16, The Hague, Nederlands. September 2016.

- 4e workshop *Recent Computational Advances in Metagenomics* (RCAM), Paris, France. September 2017.

- 50e Journées de Statistique, Palaiseau, France. May 2018.

- 5e workshop *Recent Computational Advances in Metagenomics* (RCAM), as a satellite of ECCB'18, Athens, Greece. September 2018.

- Assemblée Générale of the MIA Division of Inra, Massy-Jouy, France. May 2019.

- 6e workshop *Recent Computational Advances in Metagenomics* (RCAM), Paris, France. September 2019.

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.

- "Probability and Statistics applied to Computational Biology" in the *International Workshop in Applied Probability*, University of Compiègne, France. July 2008.

- "Probability and Statistics for Genomics" in the *International Workshop in Applied Probability*, Madrid, Spain. July 2010.

- " Applied Probability in Computational Biology" in the *International Workshop in Applied Probability*, Antalya, Turkey. June 2014.

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-Rennesson, 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. *J. Appl. Prob.* **45** 440–455.

[23] Mercier, R., Petit, M.-A., Schbath, S., Robin, S., El Karoui, M., Boccard, F. and Espeli, O. (2008). The MatP/matS site specific system organizes the Terminus region of the E. coli chromosome into a Macrodomain. *Cell.* **135** 475–485.

[24] Schbath, S., Lacroix, V. and Sagot, M.-F. (2009). Assessing the exceptionality of coloured motifs in networks. *EURASIP Journal on Bioinformatics and Systems Biology.* **ID 616234** 1–9.

[Reynaud-Bouret and Schbath (2010)] Reynaud-Bouret, P. and Schbath, S. (2010). Adaptive estimation for Hawkes' processes; Application to genome analysis. *Annals of Statistics.* **38 (5)** 2781–2822.

[25] Touzain, F., Petit, M.-A., Schbath, S. and El Karoui, M. (2011). DNA motifs that sculpt the bacterial chromosome. *Nature Reviews Microbiology.* **9** 15–26.

[26] Stefanov, V., Robin, S. and Schbath, S. (2011). Occurrence of structured motifs in random sequences: Arbitrary number of boxes. *Discrete Applied Mathematics.* doi:10.1016/j.dam.2010.12.023.

[27] Devillers, H., Chiapello, H., Schbath, S. and El Karoui, M. (2011). Robustness assessment of whole bacterial genome segmentations. *Journal of Computational Biology.* **18** 1155–1165.

[28] Devillers, H. and Schbath, S. (2012). Separating significant matches from spurious matches in DNA sequences. *Journal of Computational Biology.* **19** 1–12.

[29] Fayyaz, A., Launay, G., Schbath, S., Gibrat, J.-F. and Rodolphe, F. (2012). Statistical significance of threading scores. *Journal of Computational Biology.* **19** 13–29.

[30] SCHBATH, S., MARTIN, V., ZYTNICKI, M., FAYOLLE, J., LOUX, V. and GIBRAT, J.-F. (2012). Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. *Journal of Computational Biology*. **19** 796–813.

[31] DE PAEPE, M., HUTINET, G., SON, O., AMARIR-BOUHRAM, J., SCHBATH, S. and PETIT, M.-A. (2014). Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: The role of Rad52-like recombinases. *PLOS Genetics*. **10(3)** e1004181.

[32] MASSIP, F., SHEINMAN, M., SCHBATH, S. and ARNDT, P. (2015). How evolution of genomes is reflected in exact DNA sequence match statistics. *Molecular biology and evolution*. **32** 524–535.

[33] MASSIP, F., SHEINMAN, M., SCHBATH, S. and ARNDT, P. (2016). Comparing the statistical fate of paralogous and orthologous sequences. *Genetics*. **204** 1–7.

[34] BENOIT, G., PETERLONGO, P., MARIADASSOU, M., DREZEN, E., SCHBATH, S., LAVENIER, D. and LEMAITRE, C. (2016). Multiple comparative metagenomics using multiset k-mer counting. *PeerJ Computer Science*. **2** e94.

[35] BENOIT, G., MARIADASSOU, M., ROBIN, S., SCHBATH, S., PETERLONGO, P. and LEMAITRE, C. (2020). Simkamin: fast and resource frugal de novo comparative metagenomics. *Bioinformatics*. **36**.

[36] HUREL, J., SCHBATH, S., BOUGEARD, S., ROLLAND, M., PETRILLO, M. and F., T. (2020). Dugmo: tool for the detection of unknown genetically modified organisms with high-throughput sequencing data for pure bacterial samples. *BMC Bioinformatics*. **21**.

[37] SULTAN, I., FROMION, V., SCHBATH, S. and NICOLAS, P. (2020). Statistical modelling of bacterial promoter sequences for regulatory motif discovery with the help of transcriptome data: application to Listeria monocytogenes. *Journal of the Royal Society Interface*. **17**.

[38] AUBERT, J., SCHBATH, S. and ROBIN, S. (2021). Model-based biclustering for overdispersed count data with application in microbial ecology. *Methods in Ecology and Evolution*. **12** 1050–1061.

[39] BIZE, A., MIDOUX, C., MARIADASSOU, M., SCHBATH, S., FORTERRE, P. and DA CUNHA, V. (2021). Exploring short k-mer profiles in cells and mobile elements from archaea highlights the major influence of both the ecological niche and evolutionary history. *BMC Genomics*. **22**.

[40] MARIADASSOU, M., NOUVEL, L.-X., CONSTANT, F., MORGAVI, D., RAULT, L., BARBEY, S., HELLOIN, E., RUÉ, O., SCHBATH, S., LAUNAY, F., SANDRA, O., LEFEBVRE, R., LE LOIR, Y., GERMON, P., CITTI, C. and EVEN, S. (2023). Microbiota members from body sites of dairy cows are largely shared within individual hosts throughout lactation but sharing is limited in the herd. *Animal Microbiome*. **5** 32.

Books / Book Chapters

[41] 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`).

[42] 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.

[43] ROBIN, S., RODOLPHE, F. and SCHBATH, S. (2003). *ADN, mots et modèles*. BELIN.

[44] 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.

[45] ROBIN, S., RODOLPHE, F. and SCHBATH, S. (2005). *DNA, Words and Models*. Cambridge University Press.

[46] SCHBATH, S. and ROBIN, R. (2009). How can pattern statistics be useful for DNA motif discovery? In *Scan Statistics – Methods and Applications*, (J. Glaz, I. Pozdnyakov, and S. Wallenstein, eds.). Statistics for Industry and Technology. Birkhauser.

[47] SCHBATH, S. and HOEBEKE, M. (2011). R'MES: a tool to find motifs with a significantly unexpected frequency in biological sequences. In *Advances in genomic sequence analysis and pattern discovery*, (L. Elnitski, O. Piontkivska, and L. Welch, eds.). Science, Engineering, and Biology Informatics, vol. 7. World Scientific.

Technical Reports

[48] 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.

[49] 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.

[50] 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).

[51] SCHBATH, S. (2006). Statistics of motifs. Lecture notes for *Atelier INSERM Identification of non-coding functional regions in genome*, La Londes-les-Maures, April 27-28 (10 pages). `http://migale.jouy.inra.fr/outils/mig/rmes/atelier-inserm2006.pdf`

[52] 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`

[53] SCHAEFFER, B. and SCHBATH, S. (2010, 2016). *Initiation à R*.

Invited Conference Lectures

[54] 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 Satelite Meeting, Rostock, Germany. Aug 31 - Sept 4.

[55] 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.

[56] 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.

[57] 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.

[58] 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).

[59] 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.

[60] SCHBATH, S. (2004). Modèles statistiques et analyse de génomes. In *36èmes Journées de Statistique*. Montpellier, France. May, 22-28.

[61] SCHBATH, S. (2004). Modelling the dependence between sequence motifs. In *6th World Bernoulli Congress*. Barcelonna, Spain. July, 26-30.

[62] SCHBATH, S. (2005e). Statistical problems arising in physical mapping. In *Workshop on Stat. in Genomics and Proteomics*. Estoril, Portugal. October, 6-8.

[63] SCHBATH, S. (2005f). The statistical world of motifs on genomes. In *SemStat Summer School*. University of Warwick, UK. September, 11-13.

[64] SCHBATH, S. (2006b). Statistics of motifs. In *Atelier INSERM Identification de régions non codantes fonctionnelles dans les génomes*. La Londes-les-Maures. April, 27-28.

[65] SCHBATH, S. (2006a). Network motifs: mean and variance for the count. In *International Workshop on Applied Probability*. University of Connecticut, USA. May, 15-18.

[66] 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.

[67] SCHBATH, S. (2008). Occurrences of structured motifs along DNA sequences. In *International Workshop on Applied Probability*. Université Technologique de Compiègne, France. July, 7-10.

[68] SCHBATH, S. (2008). The statistical world of motif occurrences along dna sequences. In *Workshop Hitting, returning and matching in dynamical systems, information theory & mathematical biology*. Eindhoven, The Netherlands. November, 3-7.

[69] SCHBATH, S. (2008). Statistical analysis of biological networks; Assessing the exceptionality of network motifs. In *Approches quantitatives de la complexité biologique*. PRES UniverSud Paris, Orsay, France. May, 5-6.

[70] SCHBATH, S. (2009). Statistics of biological network motifs; A compound poisson approximation for their count in random graphs? In *Progress in Stein's method*. Singapour. January, 12-16.

[71] SCHBATH, S. (2009). R'MES: Finding exceptional motifs in sequences. In *Bioinformatics Open Source Conference*. Stockholm, Sweden. June, 28.

[72] SCHBATH, S. (2010). Statistics of biological network motifs. In *Seminar of the PhD program: Complex systems for postgenomic biology*. Cancer Research Institute, Torino, Italy. February, 24.

[73] SCHBATH, S. (2010). Statistique de mots en génomique : ce qui a été fait, ce qu'il reste à faire. In *Journées de Statistiques du Sud*. Mèze, France. June 24.

[74] SCHBATH, S. (2011). Motif-based comparison of biological networks. In *Workshop on Discrete Mathematics and Probability in Networks and Population Biology*. Singapour. May, 9-13.

[75] SCHBATH, S. (2011). Statistical models and analyses for biological networks. In *Networks research cluster workshop*. Oxford, UK. June 2.

[76] SCHBATH, S. (2012). Statistics of network motifs. In *Computational Biology Symposium*. Los Angeles, USA March 30 - April 1st.

[77] SCHBATH, S. (2012). Metabolic network comparison based on coloured motif occurrences. In *International Workshop on Applied Probability*. Jerusalem, Israel. June 11-14.

[78] SCHBATH, S. (2012b). Separating significant matches from spurious matches in DNA sequences. In *ConfÃ©rence en l'honneur des 65 ans d'Alain GuÃ©noche*. Marseille, France. October 25-26.

[79] SCHBATH, S. (2013). Mapping reads on a genomic sequence: a practical comparative analysis. In *Kick-of meeting of PF7 RADIANT project*. Manchester, UK. January 14-15.

[80] SCHBATH, S. (2014a). Assessing the enrichment significance of a position weight matrix along a DNA sequence. In *International Workshop in Applied Probability*. Antalya, Turkey. June 16-19 (Invited Lecture).

[81] SCHBATH, S. (2014b). La bioinformatique et sa sociÃ©tÃ© savante Ã l'Ã©chelle nationale. In *Rencontres METIC*. Montpellier, France. October 21.

[82] SCHBATH, S. (2015a). A la recherche de motifs statistiquement sur-représentés dans les génomes : des mots aux matrices poids-position. In *Séminaire du CMAP*. École Polytechnique, Palaiseau, France. March, 10.

[83] SCHBATH, S. (2015b). Statistics of motifs: how to deal with position-weight matrices. In *Meeting of FP7 RADIANT project*. Naples, Italy. July 2-3 (invited lecture).

[84] SCHBATH, S. (2017). Une histoire de mots innattendus et de génomes. In *Journées ALEA*. Luminy, France. March 20-24 (invited lecture).

[85] SCHBATH, S. (2018). The french Microbial Ecosystems and Meta-omics (MEM) metaprogramme from INRA. In *Hellenic Bioinformatics*. Thessalonica, Greece. November 16-18 (invited lecture).

[86] LOUX, V. and SCHBATH, S. (2018). Des microbes dans mon fromage ? In *Journées Nationales de la Science Ouverte*. Paris, France. December 6 (invited lecture).

[87] SCHBATH, S. (2019). Quels microbes pour fabriquer un nouveau jus de lupin fermenté ? Le text-mining à la rescousse ! In *VisaTM Days*. Paris, France. November 15 (invited lecture).

[88] SCHBATH, S. (2019). Text-mining: a complementary approach to bioinformatics for research in microbiology. In *DataIA Days: IA and Agronomics*. Université Paris-Saclay, France. December 4 (invited lecture).

[89] SCHBATH, S. (2021). GreenMaIAGE : initiatives pour des pratiques plus écoresponsables. In *Séminaire de l'unité GABI d'INRAE*. Jouy-en-Josas, France. May 31 (invited lecture).

[90] SCHBATH, S. (2021). GreenMaIAGE : bilan ges et initiatives pour des pratiques plus écoresponsables. In *Journée Impact carbone de la recherche et du numérique du Pôle IMABS*. Toulouse, France. September (invited lecture).

[91] SCHBATH, S. (2021). GreenMaIAGE : bilan ges et initiatives pour des pratiques plus écoresponsables. In *Séminaire du réseau des relais Développement Durable du centre INRAE de Versailles*. Versailles, France. October 22 (invited lecture).

[92] SCHBATH, S. (2021). GreenMaIAGE : bilan ges et initiatives pour des pratiques plus écoresponsables. In *Séminaire du département AgroEcoSystem d'INRAE*. Jouy-en-Josas, France. Novembre 22 (invited lecture).

[93] SCHBATH, S. (2022f). MathNum au coeur des enjeux RSE d'INRAE. In *Assemblée Générale du département MathNum d'INRAE*. Clermont-Ferrand, France. May 18 (Invited talk).

[94] SCHBATH, S. (2022a). Faire son BGES avec GES-1point5 et après ? retex de l'unité MaIAGE. In *Réunion des directeurs et directrices d'unité du centre INRAE IdF-Versailles*. Versailles, France. September 12 (Invited talk).

[95] SCHBATH, S. (2022b). Faire son BGES avec GES-1point5 et après ? retex de l'unité MaIAGE. In *Conseil du centre INRAE IdF-Versailles*. Versailles, France. September 26 (Invited talk).

[96] SCHBATH, S. (2022c). Maiage engagée dans la réduction de son empreinte carbone. In *Assemblée Générale de l'Unité de recherche HYCAR*. Antony, France. September 22 (Invited talk).

[97] SCHBATH, S. (2022d). Maiage engagée dans la réduction de son empreinte carbone. In *Atelier participatif Développement Durable et Responsabilité Sociétale de l'unité de recherche GéoSciences*. Rennes, France. November 17 (Invited talk).

[98] SCHBATH, S. (2022e). Maiage engagée dans la réduction de son empreinte carbone. In *Journées scientifiques de l'IDEEV*. Rennes, France. December 2 (Invited talk).

[99] SCHBATH, S. (2023b). Maiage engagée dans la réduction de son empreinte carbone : une analyse de la démarche de transition à destination des dus. In *Formation des directeurs et directrices d'unité, GDR Labos-1point5*. Remote, France. February 15 (Invited talk).

[100] SCHBATH, S. (2023a). Maiage engagée dans la réduction de son empreinte carbone. In *Workshop CO2ERASE, Université Dauphine*. Paris, France. June 22 (Invited talk).

Conference Lectures

[101] 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).

[102] 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).

[103] 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).

[104] 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).

[105] SCHBATH, S. (1995). Statistics of counts of words in DNA sequences. In *21st European Meeting of Statisticians*. Aarhus University, Denmark. August 21-25 (abstract).

[106] 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).

[107] 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 Medecine, Cardiff, GB. April 6-7 (abstract).

[108] SCHBATH, S. (1998). How to find exceptional words in biological sequences. In *Analyse Structurale et Fonctionnelle d'un génome*. Séminaire Algorithme et Biologie, Institut Pasteur, Paris, France. December 8-10.

[109] SCHBATH, S. (1999). A new method for protein classification based on motifs. In *Electrophoresis Forum'99*. Rouen, France. November 24-26 (abstract).

[110] 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.

[111] 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).

[112] SCHBATH, S. (2002b). Statistiques des comptages de mots dans les séquences. In *Séminaire Algorithme et Biologie*. Lyon, France. October, 1-3 (abstract).

[113] BOURGAIT, 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).

[114] 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).

[115] SCHBATH, S. (2005g). À la recherche de motifs exceptionnels dans les génomes. In *Séminaire Mathématiques Appliquées*. Université Paris V, France. April, 15 (abstract).

[116] 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).

[117] TOUZAIN, F., SCHBATH, S., DEBLED-RENNESSON, 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).

[118] SCHBATH, S. (2005c). Modeling the dependence between sequence motifs. In *Journées Algorithmique Génomique*. University of Paris XI, Orsay, France. November, 24-25 (abstract).

[119] 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).

[120] SCHBATH, S. (2005d). Network motifs: mean and variance for the count. In *CompBioNet*. Lyon, France. December, 5-7 (paper).

[121] SCHBATH, S. (2006c). Towards exceptional motifs in biological networks. In *Workshop Statistical Methods for Post-Genomics data*. INSA, Toulouse, France. March, 30-31 (abstract).

[122] 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).

[123] 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).

[124] SCHBATH, S. (2007b). Assessing the exceptionality of network motifs. In *Tripartite meeting BIOSS/INRA/Biometris*. University of Wageningen, The Netherlands (abstract).

[125] 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).

[126] 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).

[127] DEVILLERS, H., CHIAPELLO, H., EL KAROUI, M. and SCHBATH, S. (2009). How to measure the robustness of bacterial genome comparisons? In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (E. Rivals and I. Rusu, ed.), Nantes, France (**selected long paper**, 25–30).

[128] DEVILLERS, H., CHIAPELLO, H., SCHBATH, S. and EL KAROUI, M. (2010). Assessing the robustness of complete bacterial genome segmentations. In *RECOMB-CG 2010*, (E. Tannier, ed.). Lecture Notes in Bioinformatics. **6398** 173-187 (**selected long paper**).

[129] KOSKAS, M., GRASSEAU, G., BIRMELÉ, E., SCHBATH, S. and ROBIN, S. (2011). Nemo: Fast count of network motifs. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (E. Barillot, C. Froidevaux, and E. Rocha, ed.), Paris, France (**selected long paper**, 53–60).

[130] FAYYAZ, A., LAUNAY, G., SCHBATH, S., GIBRAT, J.-F. and RODOLPHE, F. (2011). How significant is a threading score? In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (E. Barillot, C. Froidevaux, and E. Rocha, ed.), Paris, France (**selected long paper**, 27–34).

[131] KOSKAS, M., GRASSEAU, G., BIRMELÉ, E., SCHBATH, S. and ROBIN, S. (2011). Nemo: Fast count of network motifs. In *Modèles et Analyse de Réseaux : Approches Mathématiques et Informatique (MARAMI)*. Grenoble, France. October 19-21.

[132] SCHBATH, S., MARTIN, V., ZYTNICKI, M. , FAYOLLE, J., LOUX, V. and GIBRAT J.-F. (2012). Mapping reads on a genomic sequence: a practical comparative analysis. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (Coste, F. and Tagu, D., ed.), Rennes, France (**selected long paper**, 183–190).

[133] DUMAZERT, J., STEPHAN, J.-Y., PETIT, M.-A. and SCHBATH, S. (2013). Assessing the enrichment significance of a possition weight matrix (PWM) along a DNA sequence. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (C. Gaspin and e. Lindley, N., ed.), Toulouse, France (**selected long paper**, 25–34).

[134] AUBERT, J., ROBIN, S. and SCHBATH, S. (2014). Metagenomics data analysis using a latent block model: application to plant-microbial communities interactions in the rhizosphere. In *9th European Conference on Mathematical and Theoretical Biology (ECMTB)*. Gothenburg, Sweden. June 15-19 (résumé).

[135] AUBERT, J., SCHBATH, S., MOUGEL, C. and ROBIN, S. (20). Latent block model for ecological abundance data. In *30th European Meeting of Statisticians*. Amsterdam, Nederlands. July 6-10 (résumé).

[136] AUBERT, J., SCHBATH, S. and ROBIN, S. (2016a). Latent block model for metagenomic data. In *17e Journées Ouvertes Biologie Informatique Mathématiques (JOBIM 2016)*. Lyon, France. June 28-30 (**Selected long paper**).

[137] AUBERT, J., SCHBATH, S. and ROBIN, S. (2016b). Latent block model for metagenomic data. In *Workshop on Recent Computational Advances in Metagenomics (RCAM)*. The Hague, Nederlands. September 4 (résumé).

[138] BENOIT, G., PETERLONGO, P., MARIADASSOU, M., DREZEN, E., SCHBATH, S., LAVENIER, D. and LEMAITRE, C. (2017). Simka: large scale de novo comparative metagenomics. In *Journées Ouvertes Biologie Informatique Mathématiques (JOBIM)*, (C. Lhoussaine and Touzet, H., ed.), 3–5, Lille, France. (**Selected long paper**).

[139] SCHBATH, S. (2018). À la recherche de mots exceptionnels dans les génomes. In *K'fêt des sciences du collège de La Guyonnerie*. Bures-sur-Yvette, France. March 9.

[140] SULTAN, I., SCHBATH, S. and NICOLAS, P. (2018). Statistical modelling of bacterial promoter sequences for regulatory motif discovery with the help of transcription profiles. In *50e Journées de Statistique de la SFdS (JdS'2018)*. Palaiseau, France. May 28 - June 1st (résumé).

[141] SULTAN, I., SCHBATH, S. and NICOLAS, P. (2018). Statistical modeling of bacterial promoter sequences for regulatory motif discovery using expression data. In *Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM)*. Marseille, France. July 3-6 (**Selected long paper**).

[142] GOUTORBE, B., ABRAHAM, A.-L., MARIADASSOU, M., PLAUZOLLES, A., BIDAUT, G., HALFON, P. and SCHBATH, S. (2021). Shallow shotgun metagenomics as a cost-effective and accurate alternative to wgs for taxonomic profiling and clinical diagnosis. In *Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM)*. Paris, France. July 6-9 (**Selected long paper**).

STUDENTS

Ph.D. 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. (Feb 2011). *Statistics of complete bacterial genome comparisons*, University of Evry. Co-supevision with M. El Karoui.

MASSIP, F. (October 2015). *Mathematical models for chromosome evolution*, University of Paris-Saclay. Co-supevision with P. Arndt.

AUBERT, J. (February 2017). *Statistical methods for high-throughput omic data analysis*, University of Paris-Saclay. Co-supevision with S. Robin.

SULTAN, I. (June 2019). *Statistical modeling of bacterial promoter sequences for regulatory motif discovery*, University of Paris-Saclay. Co-supevision with P. Nicolas.

GOUTORBE, G. (started on september 2019). *Developement and application of an accurate and efficient method to study human microbiome for medical purpose*, University of Paris-Saclay. Co-supevision with P. Halfon and G. Bidaut.

Post-doctoral fellows

AFSHIN FAYYAZ MOVAGHAR (12 months, 2008-2009). *Statistical significance of threading scores.*

JULIEN FAYOLLE (18 months, 2009-2011). *Comparison of mapping tools for short reads.*

Engineer fellows

JÉRÔME COMPAIN (18 months, 2012-2014). *Comparison of mapping tools for short reads.*

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 statistisque 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).

Thabet, H. (2007). *Analyse statistique du nombre de mots communs à deux génomes.* Master's thesis, Modélisation Aléatoire - Universtité Paris 7. (5 months).

Ariste, O. (2010). *Evolution du motif Chi chez Escherichia coli et Staphylococcus aureus.* Bachelor's thesis, Licence professionnelle - IUT de Perpignan. (5 months).

Lacour, H. (2012). *Comparaison de méthodes de classification de séquences d'ADN et application à des données métagénomiques.* Bachelor's thesis, Institut de Statistique de Paris. (4 months).

BELISSEN, V., CABRAL-MORAES, J.-F., DARTEVELLE, B., DUMAZERT, J., FAUGEROUX, R. and STEPHAN, J.-Y. (2012). *Analyse de génome par une approche statistique, l'utilisation des matrices poids-position.* Projet Scientifique Collectif, 2nd year of Ecole Polytechnique. (8 months).

PRIVAT, M. (2014). Logiques cis-régulatrices dans le développement du cerveau moyen. Master's thesis, ENS Lyon - L3. (2 months).

AHMED ZAID, L. (2017). Analyse du lien entre puissance métabolique et entropie du génome chez les microorganismes. Master's thesis, University of Bordeaux. (4 months).

LAO, J. (2017). Comparaison de méthodes statistiques d'inférence de réseaux de co-occurences au sein d'écosystèmes microbiens à partir de données métagénomiques. Master's thesis, University Paris-Diderot. (6 months).

DULAC, G. (2020). Taxonomic classification of hyper variable regions with machine learning. Research paper. Master Data science for business. X-HEC.

December 23, 2021.