

COUR DES COMPTES

ANNEXE 3 DU RAPPORT N°139

MAI 2018

ÉVALUATION

BOURSES ET PRÊTS D'ÉTUDES

ANNEXE MÉTHODOLOGIQUE 3

**ANALYSES DE SÉQUENCES :
TRAJECTOIRES « ATYPIQUES » DES ÉLÈVES
DU SECONDAIRE II ET DU TERTIAIRE**



COUR DES COMPTES
RÉPUBLIQUE ET CANTON DE GENÈVE

Table des matières

1.	Introduction	3
2.	Le type collège (N=154)	4
3.	Le type formation professionnelle (N=244).....	7
4.	Le type école de culture générale (N=105)	10
5.	Le type interruption de formation (N=50).....	13
6.	Le type haute école spécialisée (N=54)	16
7.	Le type université (N=111)	19
8.	Le type interruption de formation (N=62).....	22
9.	Bibliographie.....	25

1. INTRODUCTION

Pour ces analyses, nous avons pris le parti de mobiliser une méthode statistique multivariée issue de la biologie moléculaire (Watermann, 1995 ; Delcher et al., 1999) : l'analyse d'appariement optimal. L'objectif réside moins ici à faire ressortir des logiques explicatives causales mais plutôt à mettre en lumière des configurations d'éléments spécifiques qui peuvent se constituer à travers l'élaboration d'idéaux-types. L'appariement optimal procède par regroupement en estimant la ressemblance entre différentes séquences. L'unité d'analyse utilisée dans ce rapport est l'individu représenté par une séquence d'états d'une durée d'un mois (sous-échantillon « secondaire II ») ou d'une année (sous échantillon « tertiaire »). Une procédure itérative estime la distance minimale séparant chaque paire de séquences (ou deux trajectoires). Le calcul de la distance qui sépare deux séquences a pour objectif de déterminer le coût des transformations à opérer afin de les rendre semblables. Les coûts sont de deux ordres : insertion/suppression (*indel*)¹ d'un élément de la séquence ou substitution d'un élément par un autre.

Le résultat de l'appariement optimal est une série de scores de similarité qui décrit l'ensemble des séquences individuelles prises deux à deux. Une analyse en clusters est ensuite appliquée à la matrice carrée symétrique formée par les scores de similarité. Pour ce faire, nous utilisons la méthode Ward qui est très largement admise dans les analyses utilisant l'appariement optimal et publiée à ce jour (Blanchard, Bühlmann et Gauthier, 2014). Il devient alors aisé d'analyser et de comparer les regroupements ainsi créés et de les représenter graphiquement. De plus, et sur la base des résultats de l'analyse de séquences, nous avons mesuré l'évolution du rapport à la formation de manière transversale. Pour ce faire, nous mobilisons la notion d'entropie, issue de la théorie de l'information et développée par Claude Shannon (1948), qui nous permet de décrire comment la diversité des statuts dans l'emploi évolue à travers le temps. Dans notre cas, une entropie faible traduira une faible diversité des statuts occupés, alors qu'une entropie forte indiquera une forte diversité des statuts entre individus.

L'opérationnalisation de cette analyse été réalisée à l'aide du logiciel R. L'analyse et la visualisation des séquences, quant à elles, sur le *package TraMineR* (Gabadinho et al., 2009).

Nous présentons ci-après les résultats graphiques pour les sous-échantillons des élèves du secondaire II et en formation tertiaire. Nous précisons encore que ces analyses n'ont porté que sur les parcours que nous avons qualifiés d'« atypiques ». En effet, l'homogénéité des trajectoires dites « typiques » constitue *per se* des modèles que nous avons nommés « G1-TYP » dans les deux sous-échantillons :

- **élèves du secondaire II (N=829),**
- **élèves du tertiaire (N=326).**

Les analyses de séquences ainsi que la représentation des courbes d'entropie transversale portent sur une période de cinq ans (septembre 2012 à juin 2017) séquencée mensuellement pour les élèves du secondaire II et annuellement pour les étudiants du tertiaire.

¹ Il s'agit des coûts d'ajout/suppression d'éléments dans la séquence : plus leur coût est élevé, plus les coûts attribués grâce à la matrice structureront la typologie dès lors que le passage d'une séquence à une autre par une succession de substitutions coûtera moins cher qu'une succession d'ajouts/suppressions. Nous soulignons encore que « l'insertion ou la suppression d'un élément induisent en contrepartie une altération de la structure temporelle des séquences comparées et elles déforment le temps. À l'inverse, les opérations de substitution préservent la structure temporelle des séquences, puisqu'elles comparent des situations situées au même moment de la séquence, mais altèrent l'enchaînement des événements » (Robette, 2011, p. 38).

2. LE TYPE COLLÈGE (N=154)

Figure 1. Séquences individuelles
G2-COLL



Figure 2. Séquences transversales

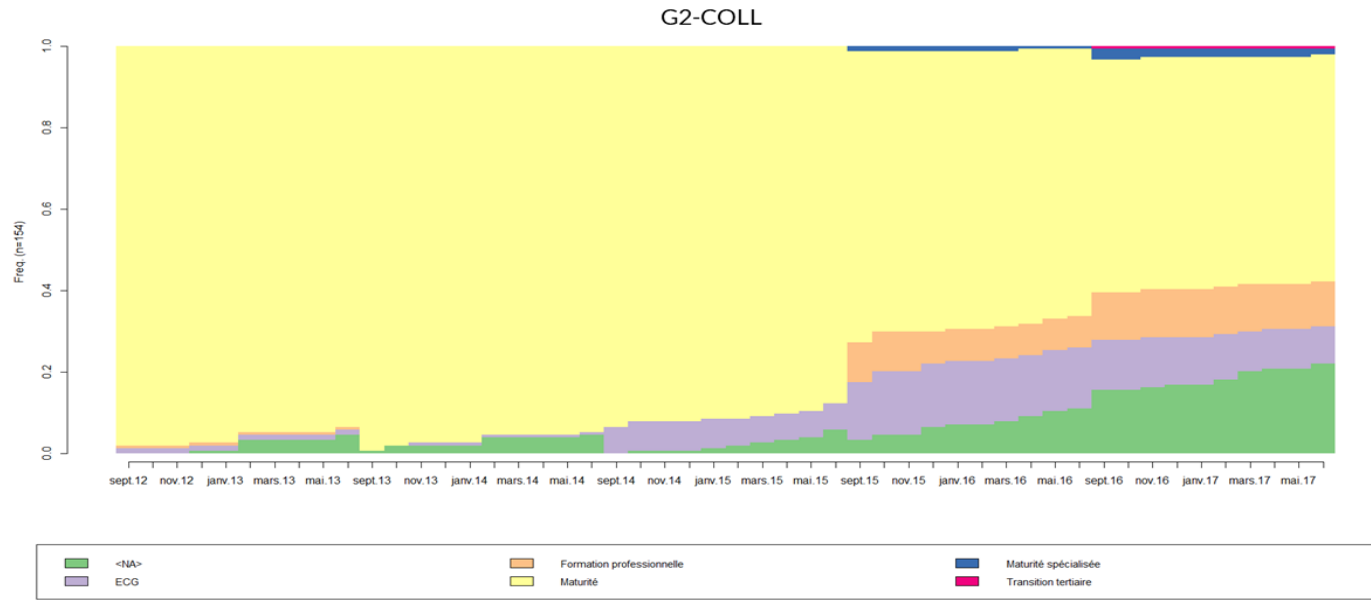


Figure 3. Mesure de diversité : entropie transversale

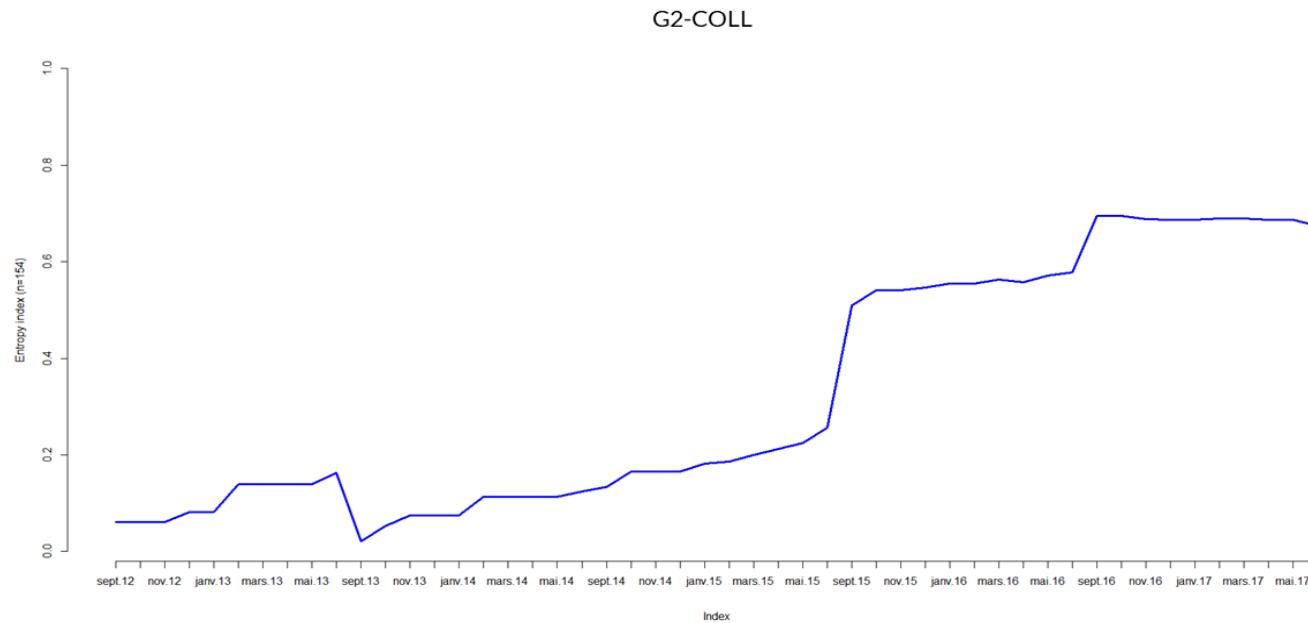
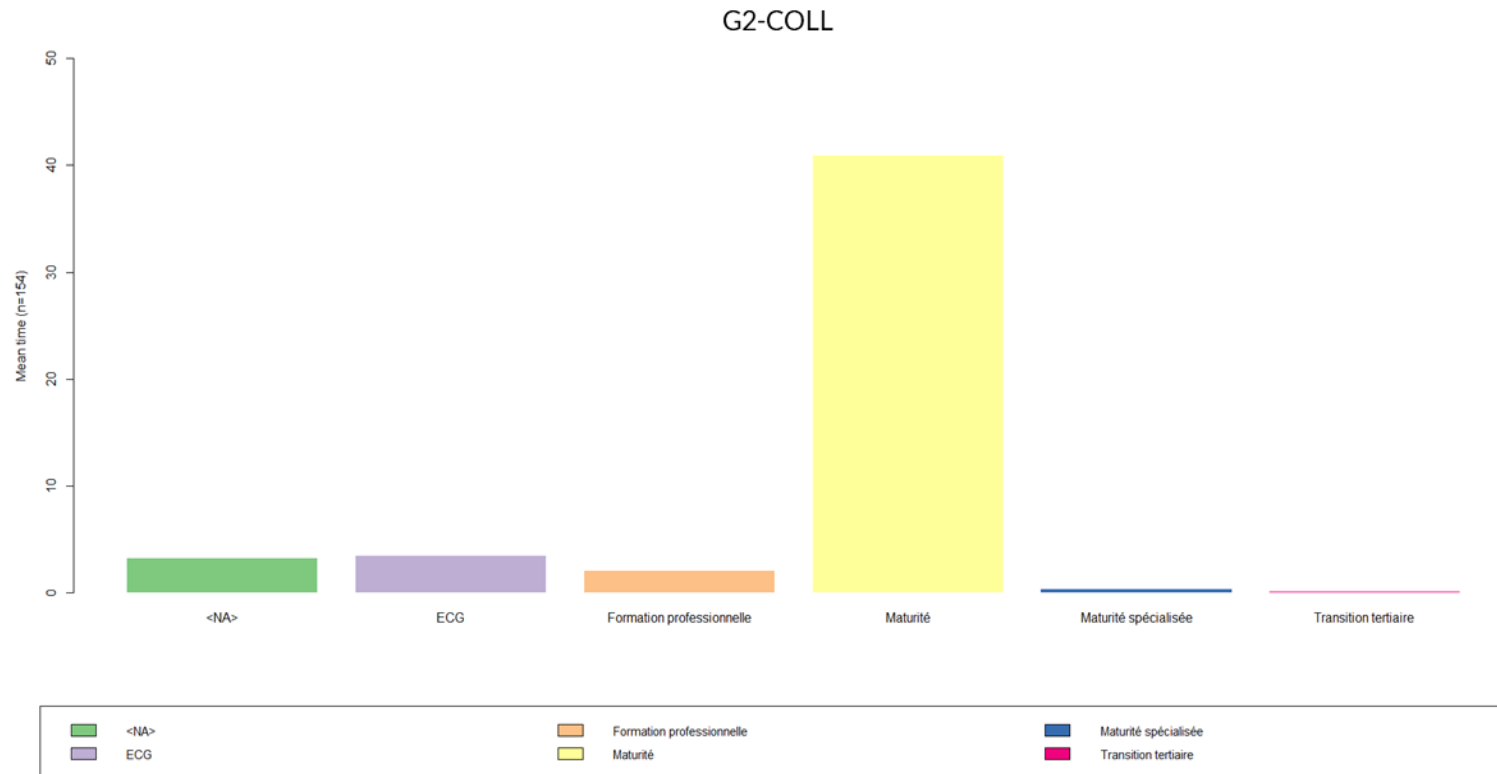


Figure 4. Temps moyen passé dans chaque état (unité de mesure = mois)



3. LE TYPE FORMATION PROFESSIONNELLE (N=244)

Figure 5. Séquences individuelles
G3-FP

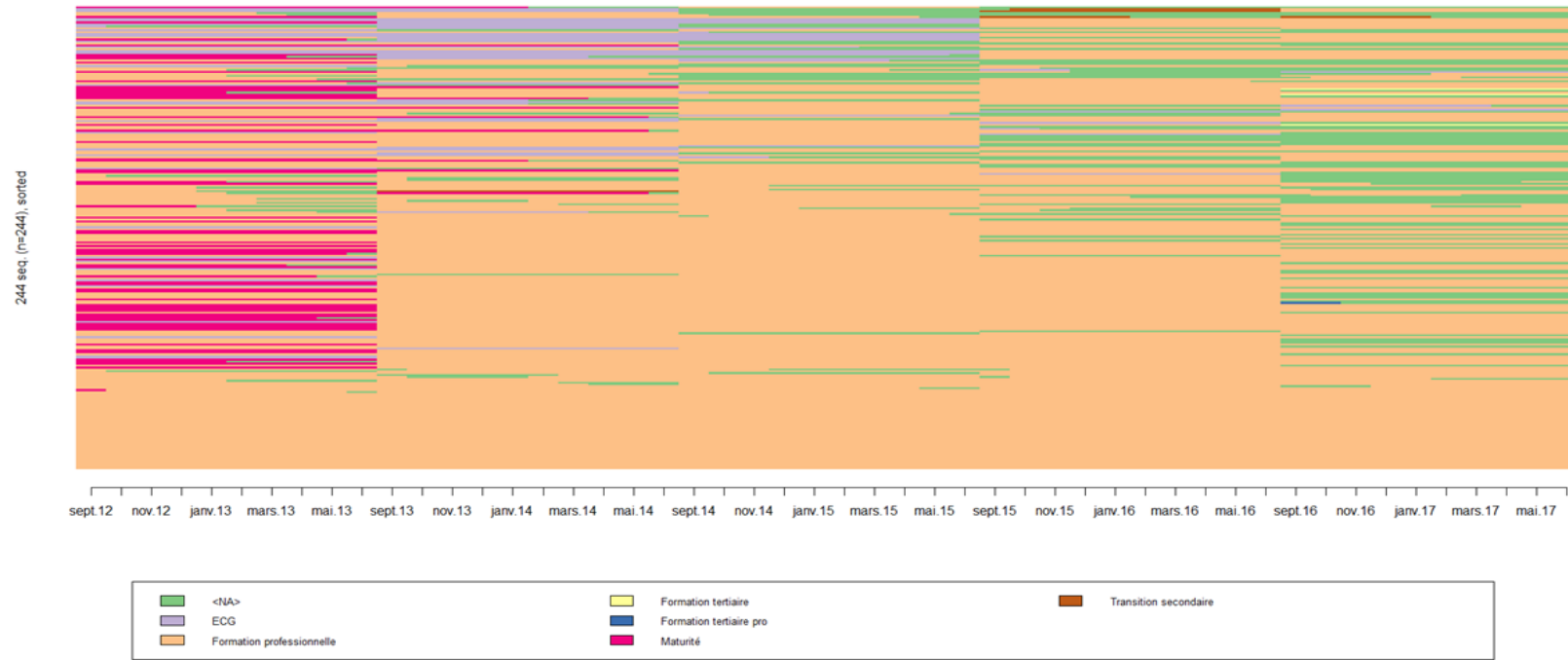


Figure 6. Séquences transversales

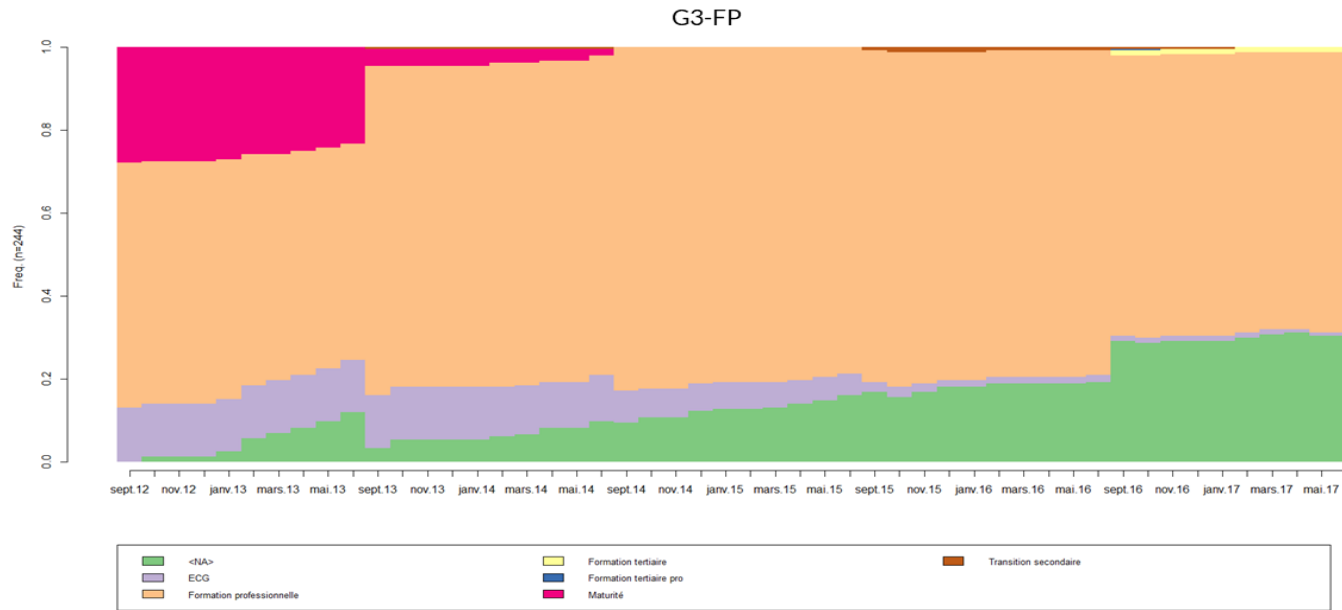


Figure 7. Mesure de diversité : entropie transversale

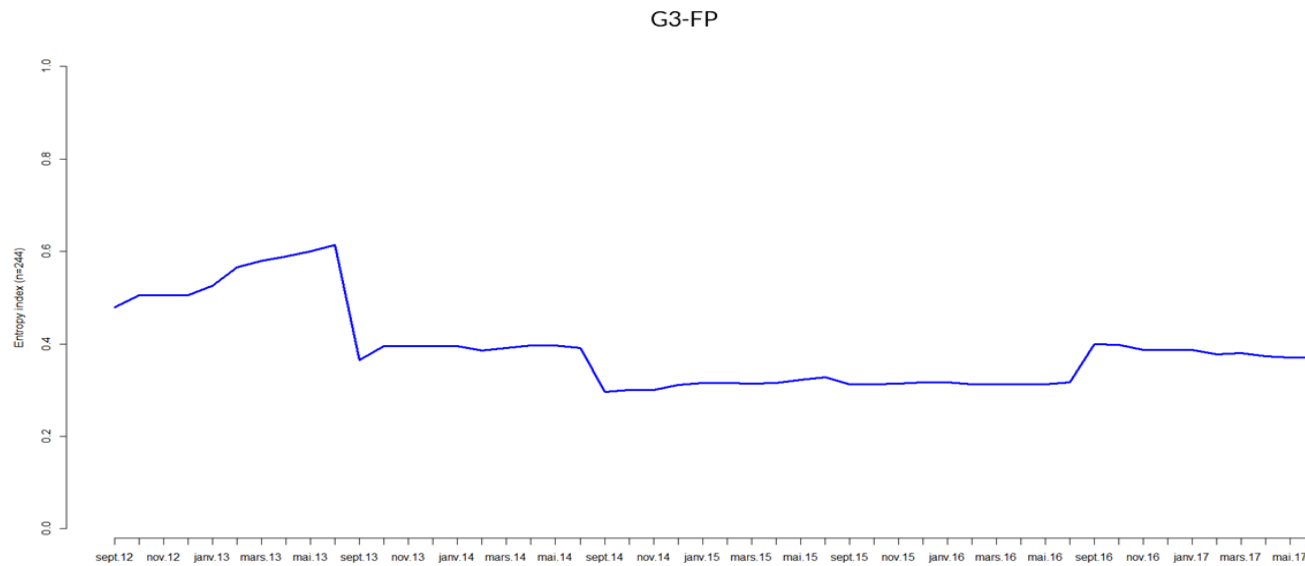
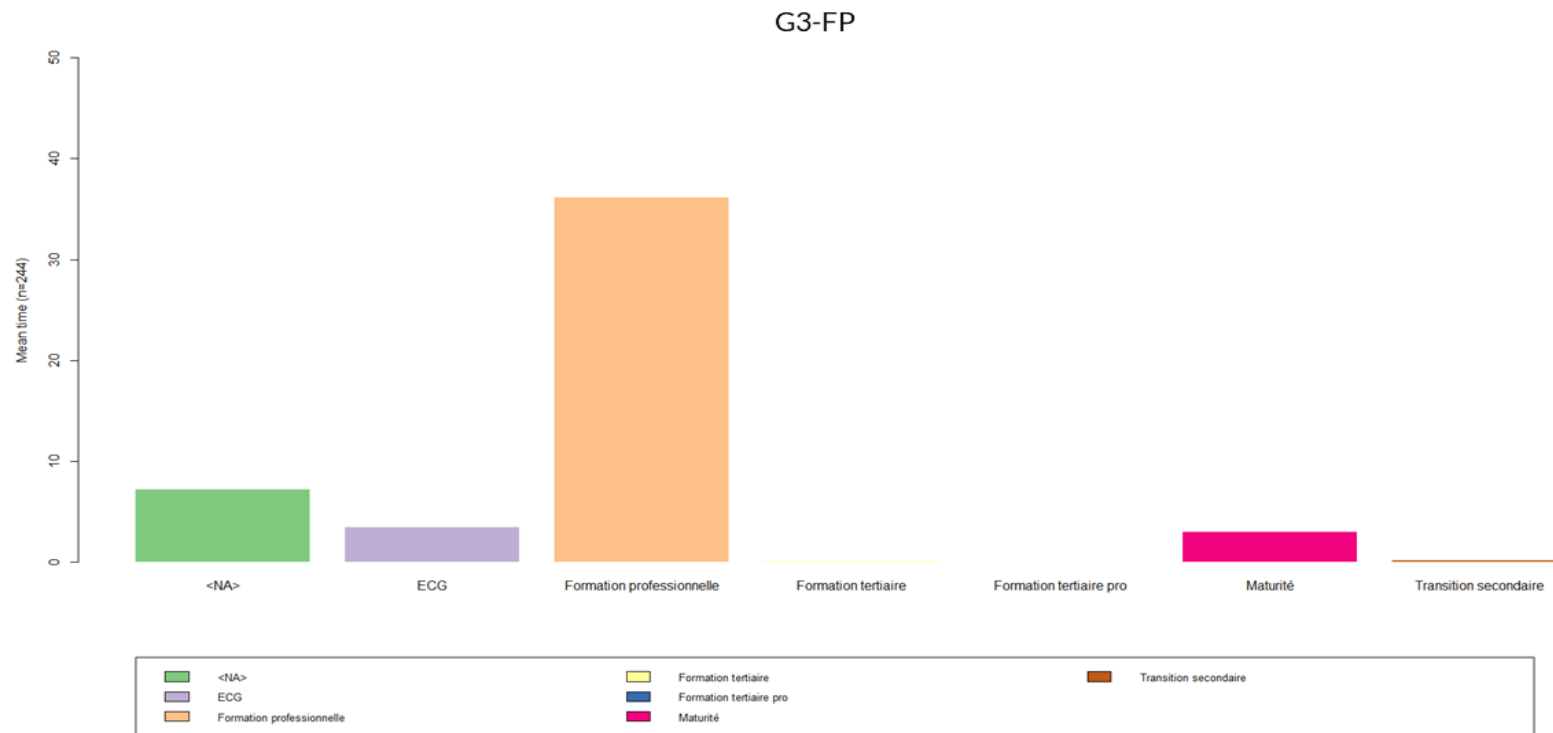


Figure 8. Temps moyen passé dans chaque état (unité de mesure = mois)



4. LE TYPE ÉCOLE DE CULTURE GÉNÉRALE (N=105)

Figure 9. Séquences individuelles
G4-ECG

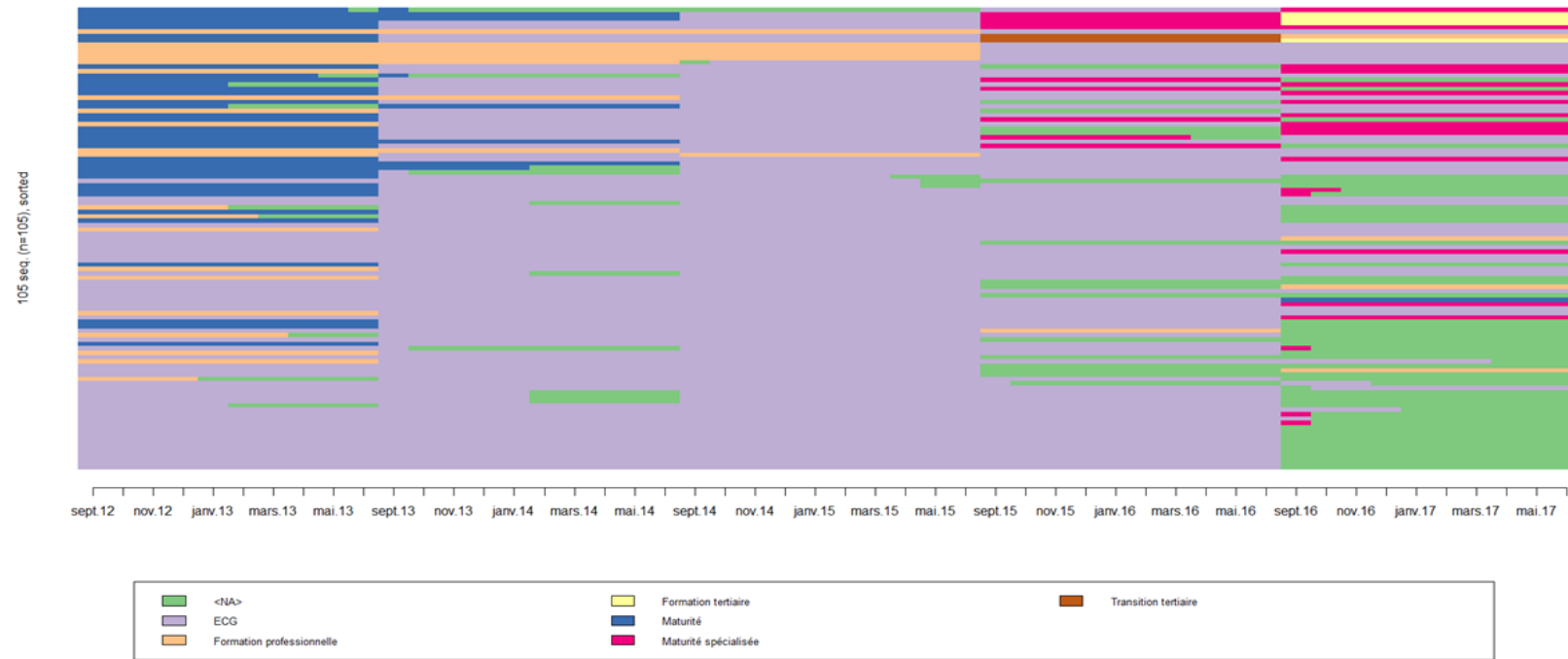


Figure 10. Séquences transversales

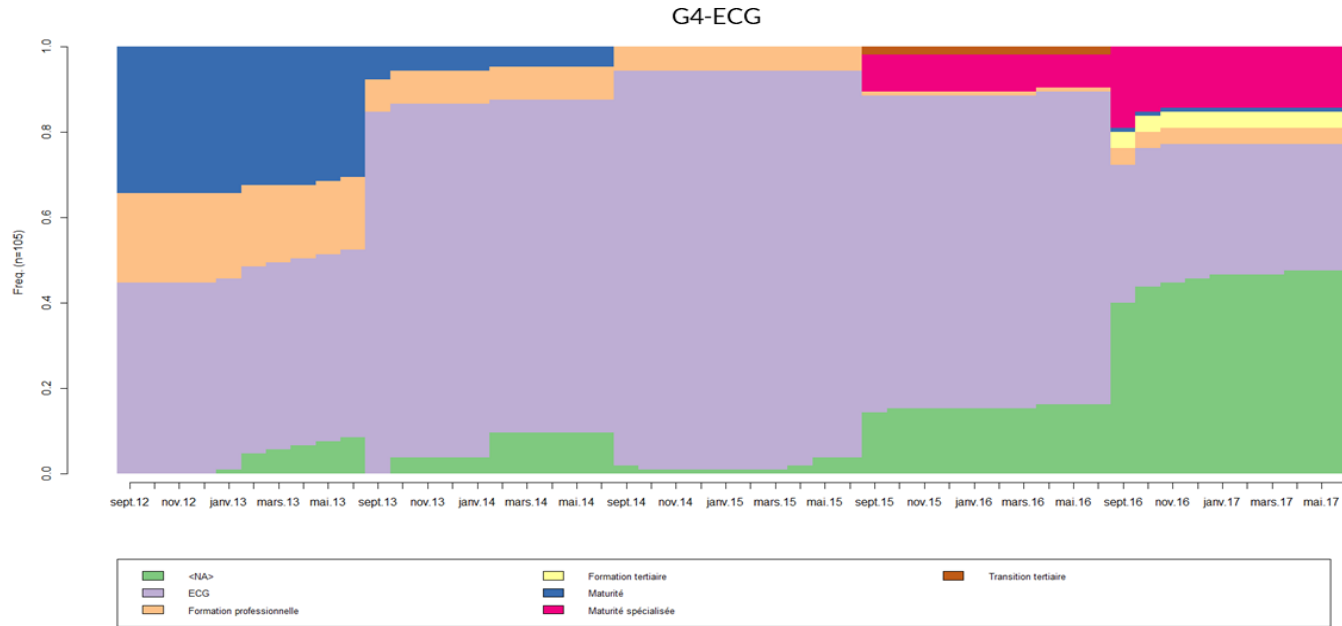


Figure 11. Mesure de diversité : entropie transversale

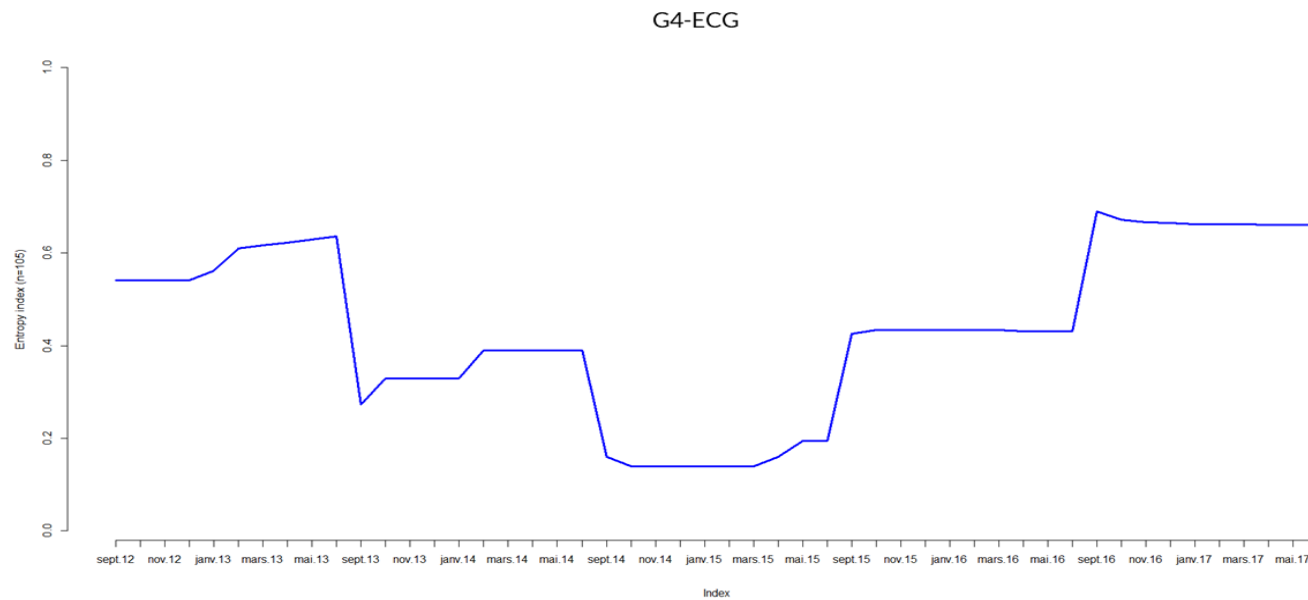
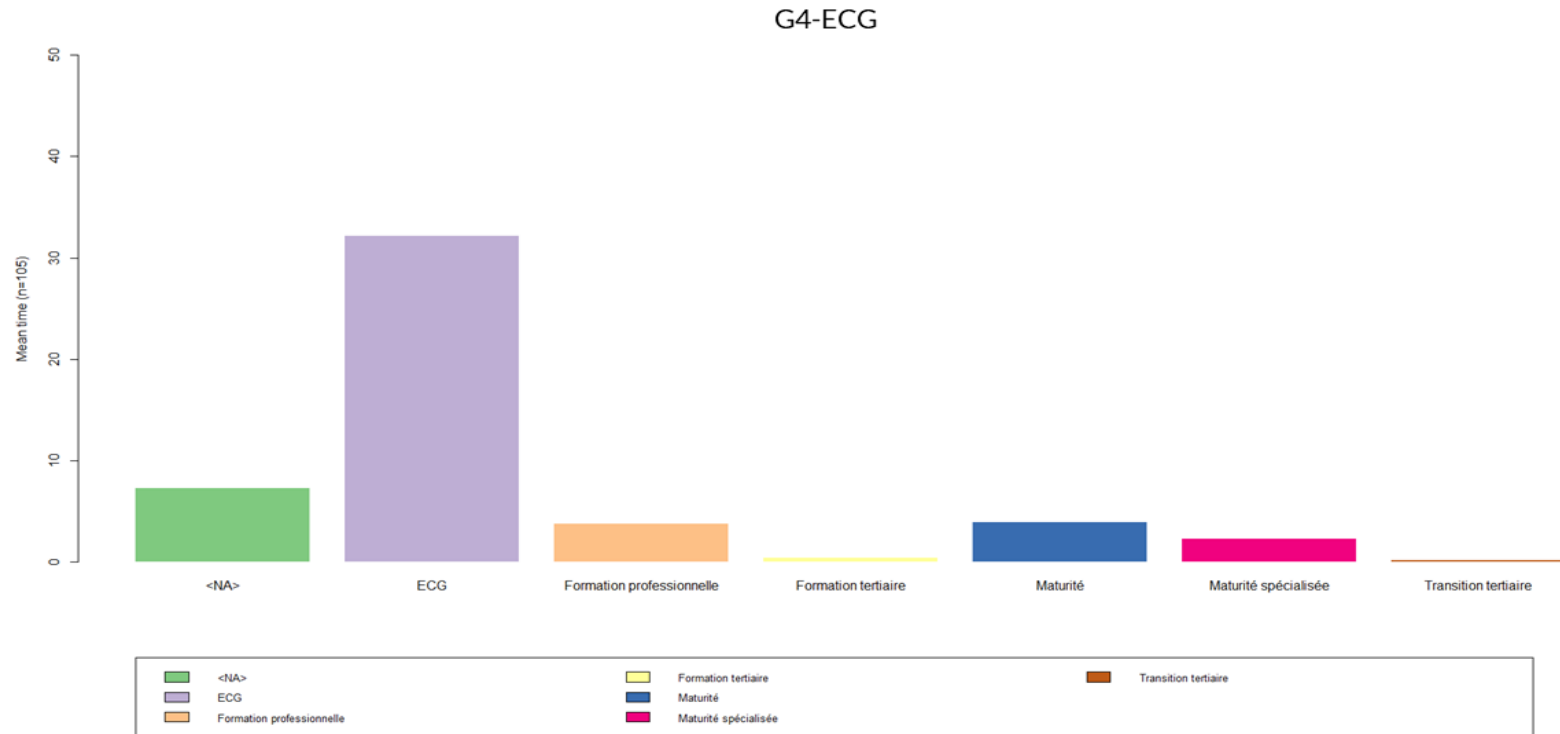


Figure 12. Temps moyen passé dans chaque état (unité de mesure = mois)



5. LE TYPE INTERRUPTION DE FORMATION (N=50)

Figure 13. Séquences individuelles

G5_NA

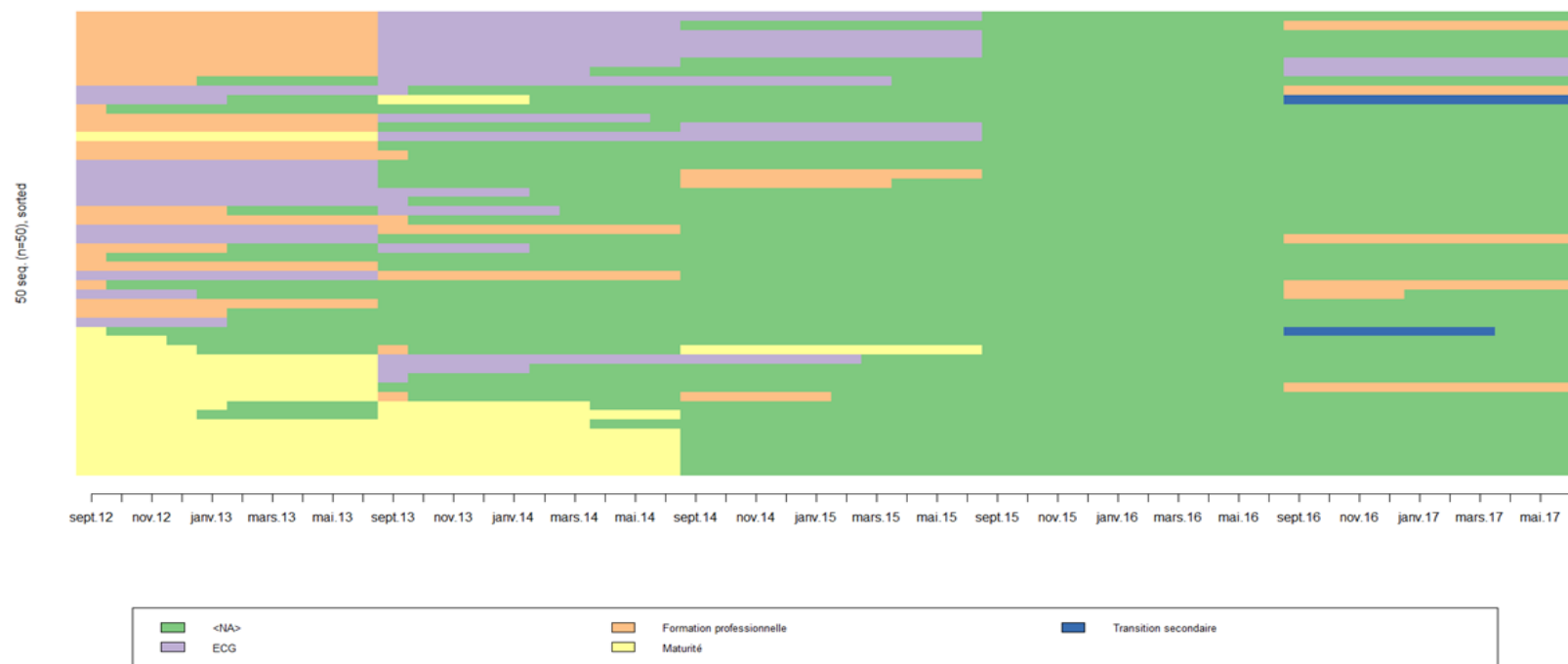


Figure 14. Séquences transversales

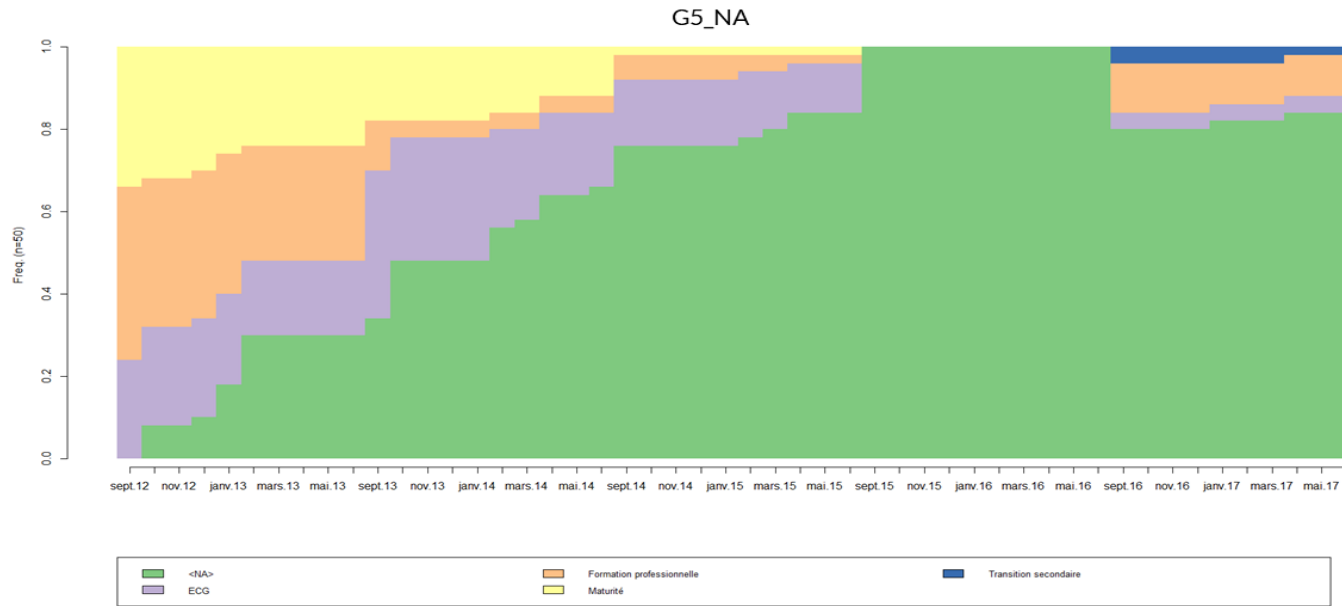


Figure 15. Mesure de diversité : entropie transversale

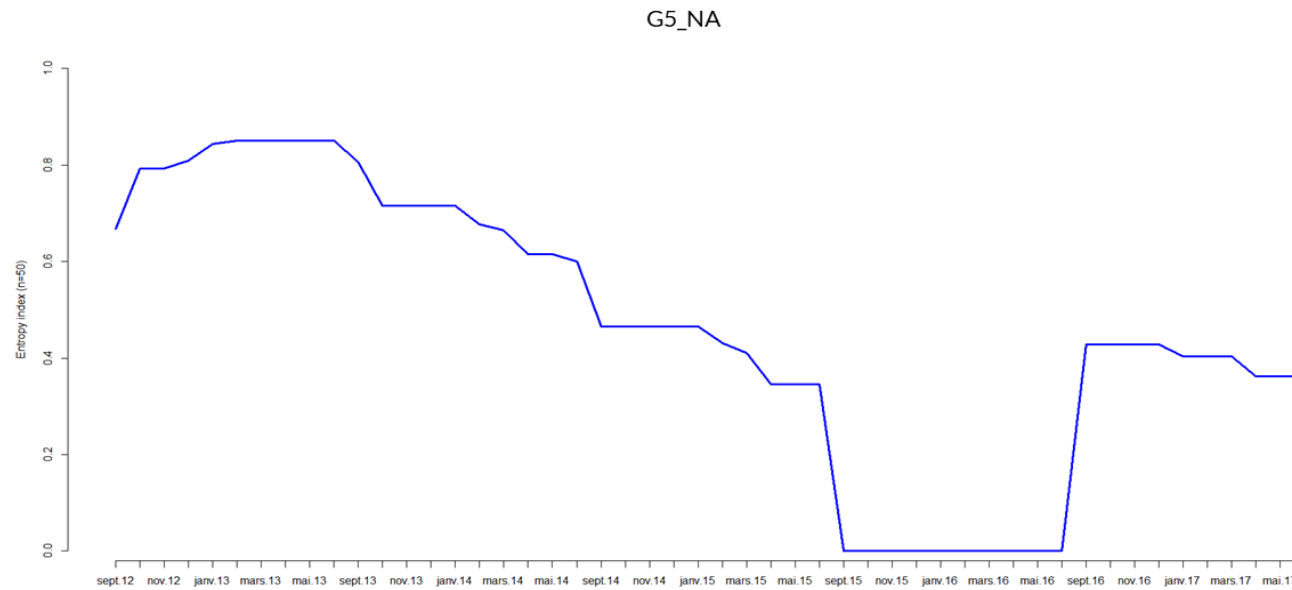
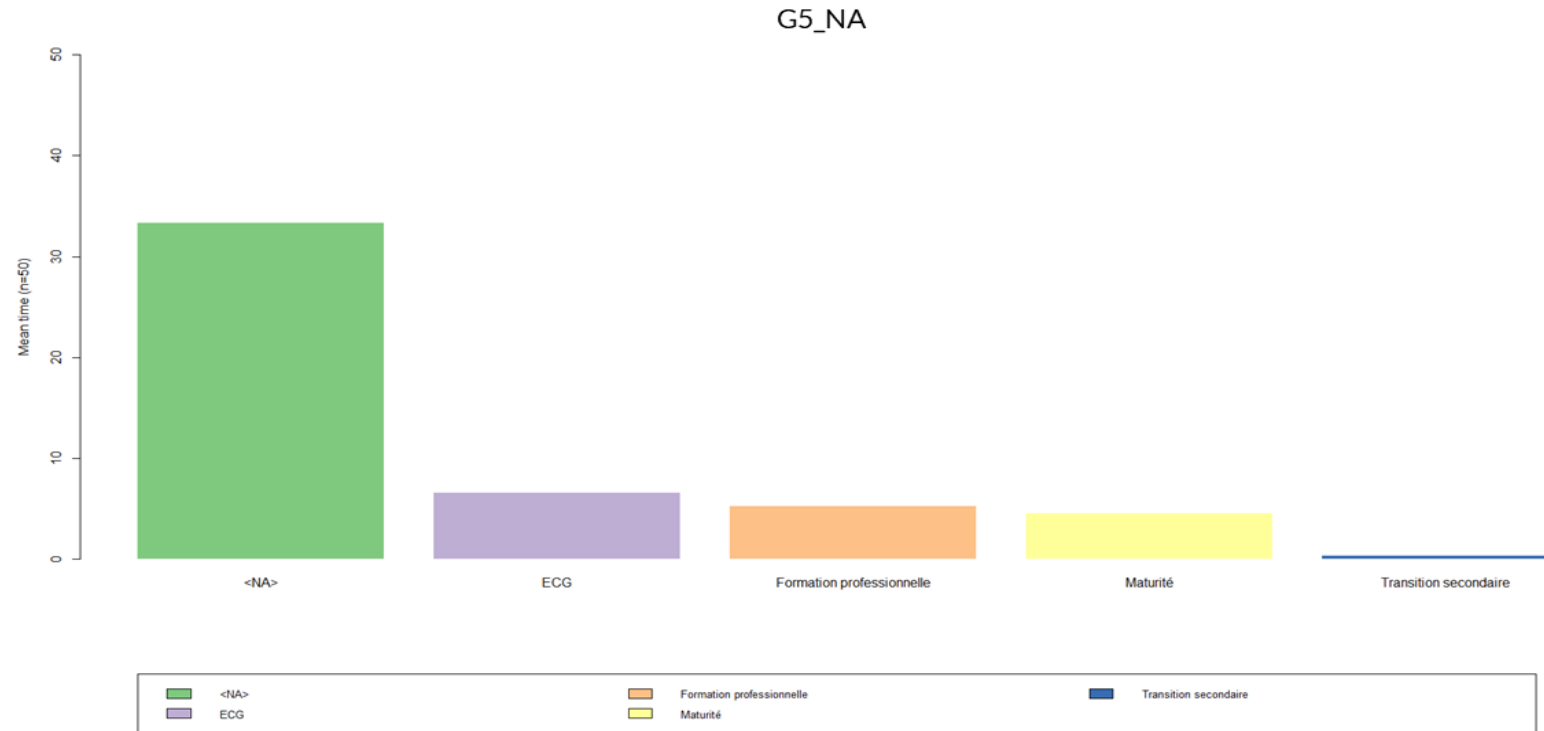


Figure 16. Temps moyen passé dans chaque état (unité de mesure = mois)



6. LE TYPE HAUTE ÉCOLE SPÉCIALISÉE (N=54)

Figure 17. Séquences individuelles

G2-HES

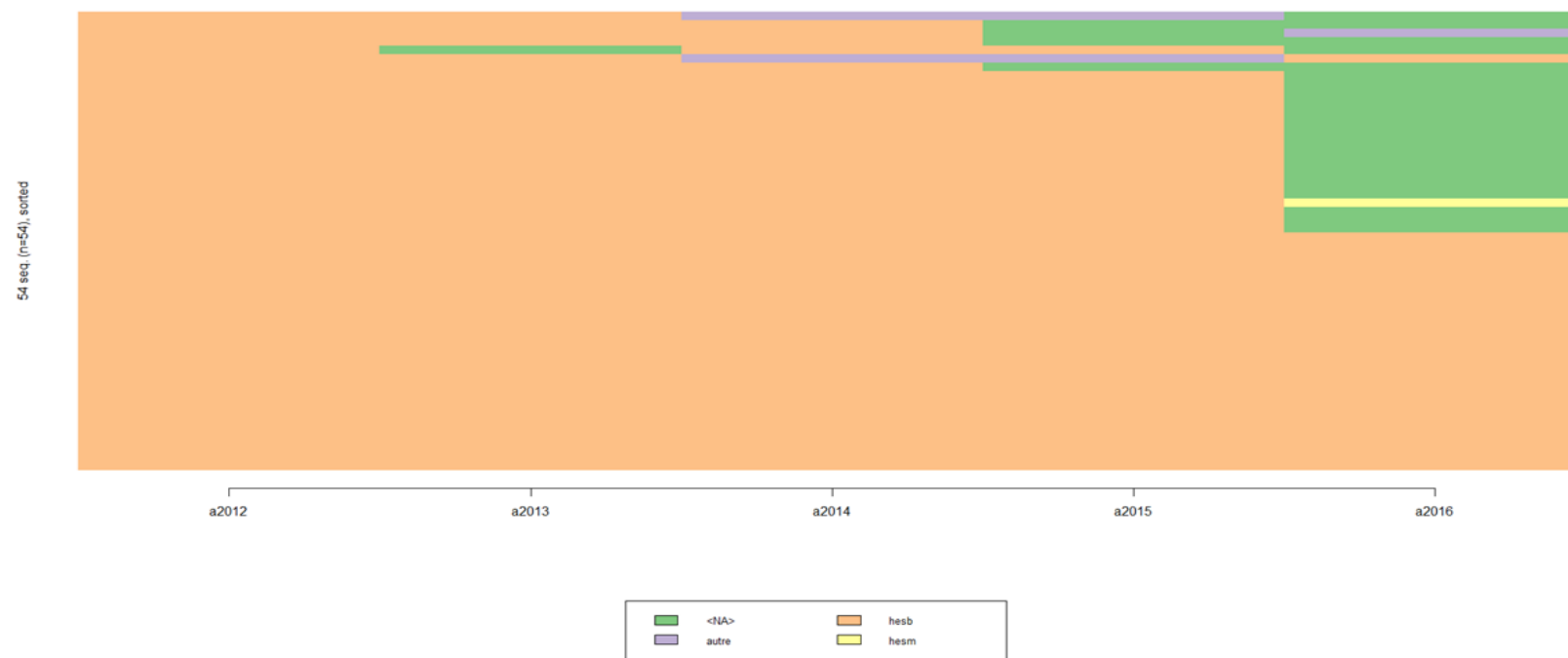


Figure 18. Séquences transversales

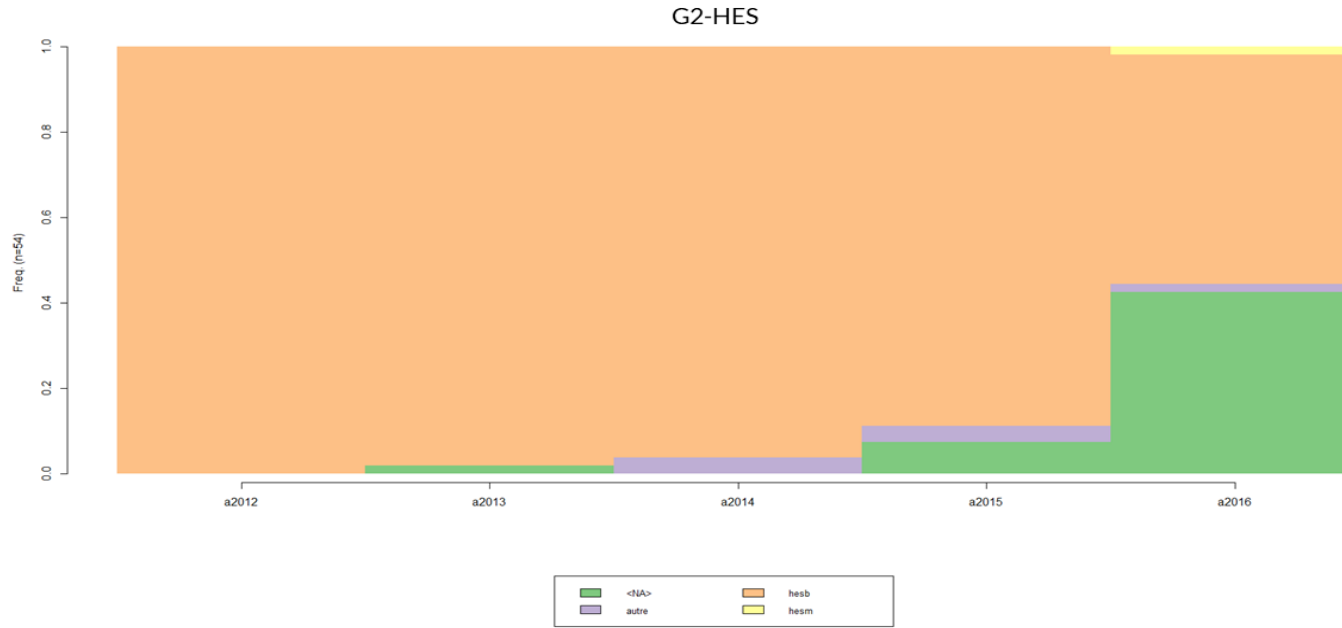


Figure 19. Mesure de diversité : entropie transversale

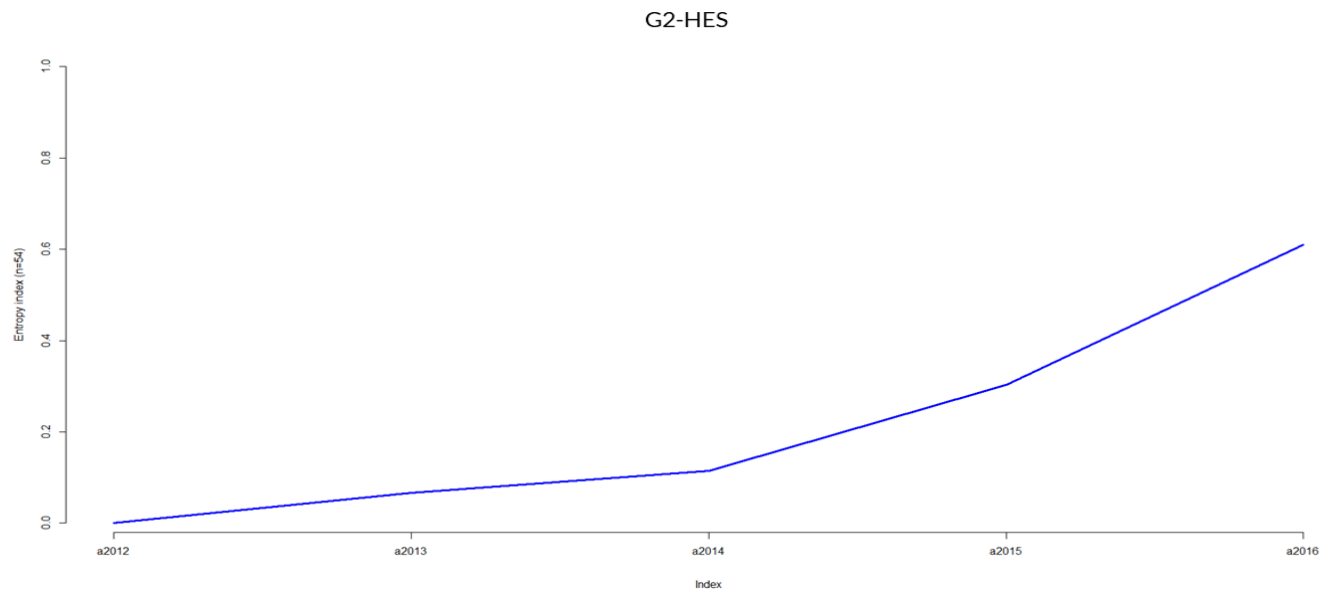
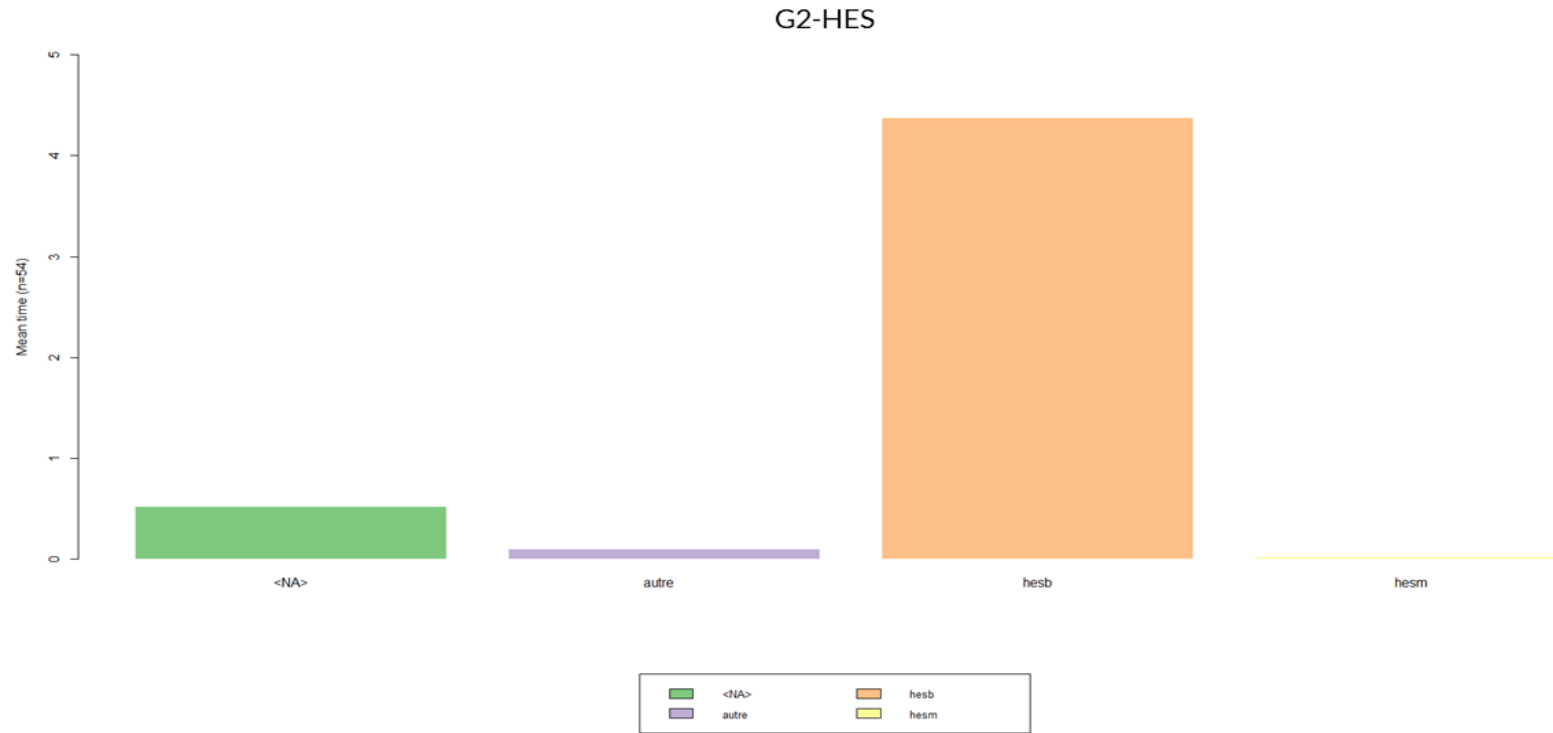


Figure 20. Temps moyen passé dans chaque état (unité de mesure = année)



7. LE TYPE UNIVERSITÉ (N=111)

Figure 21. Séquences individuelles
G3-UNI

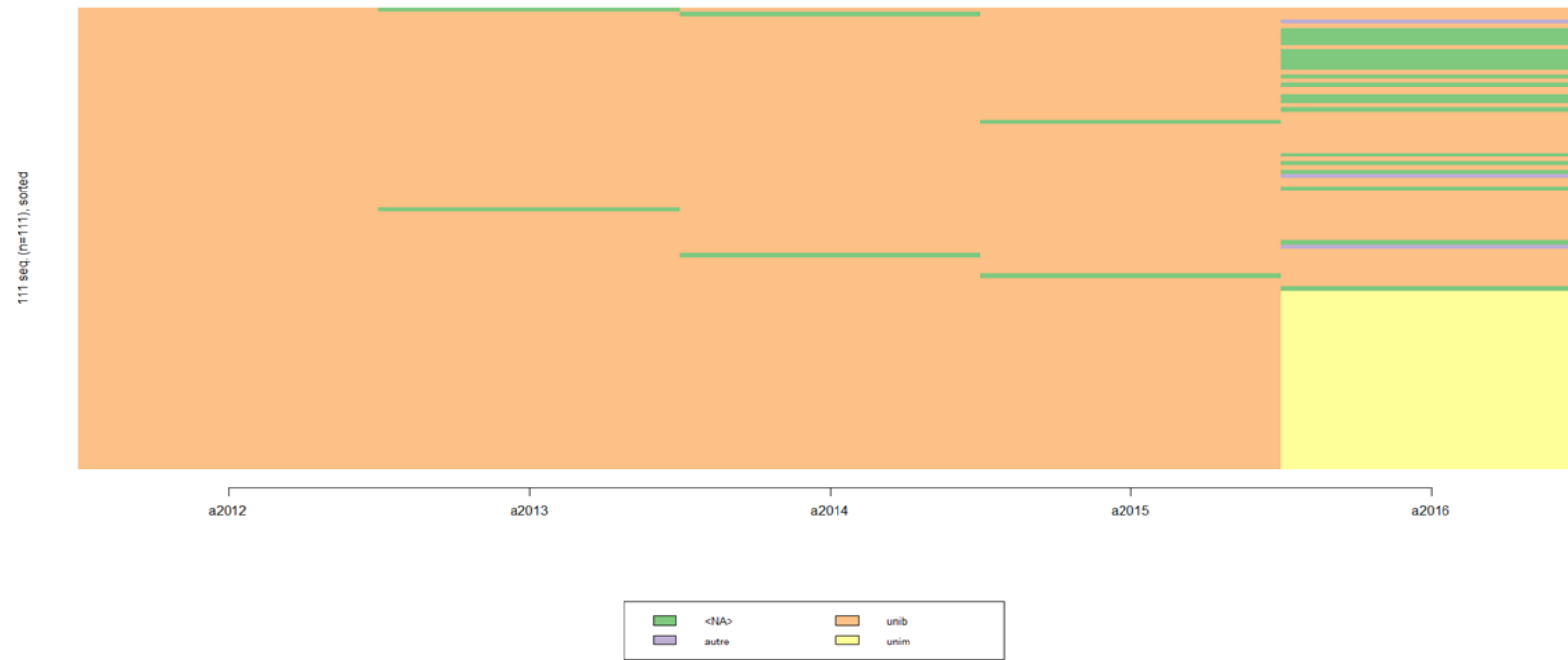


Figure 22. Séquences transversales

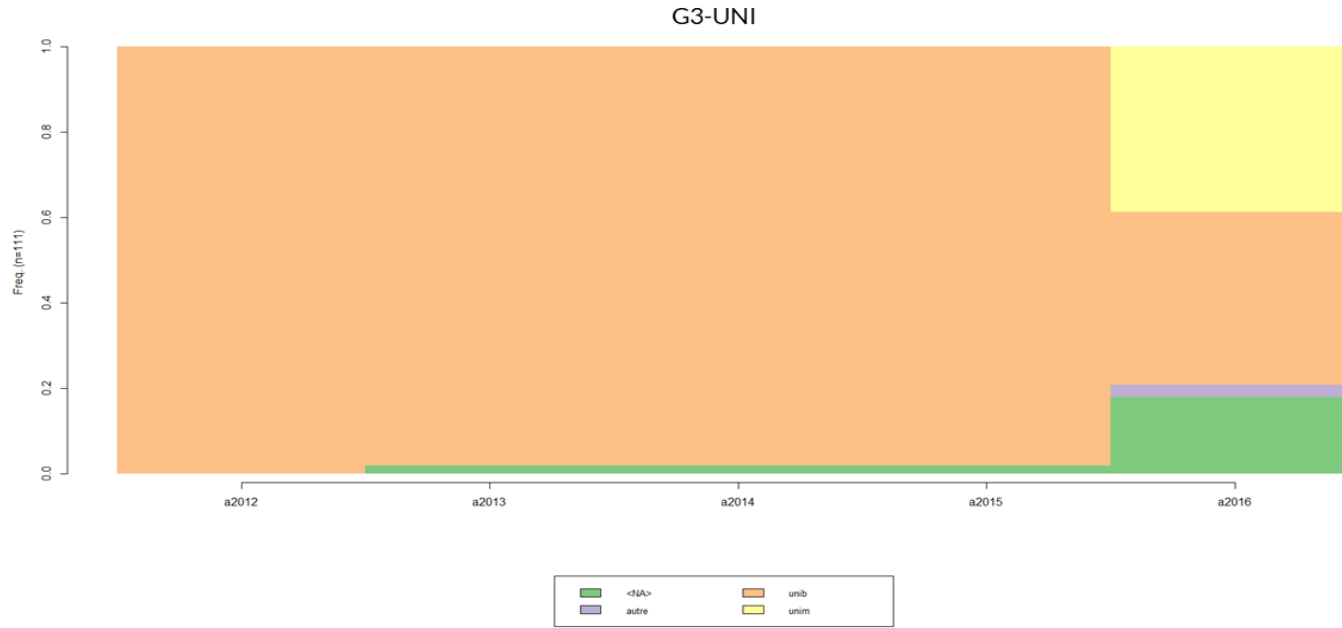


Figure 23. Mesure de diversité : entropie transversale

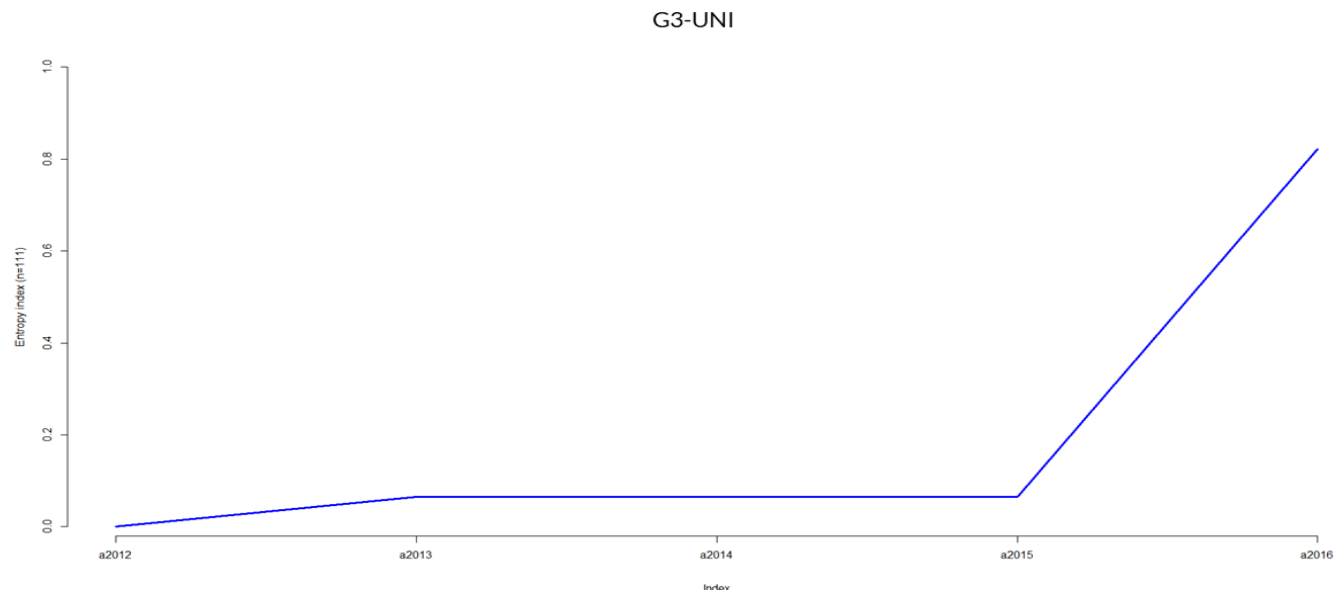
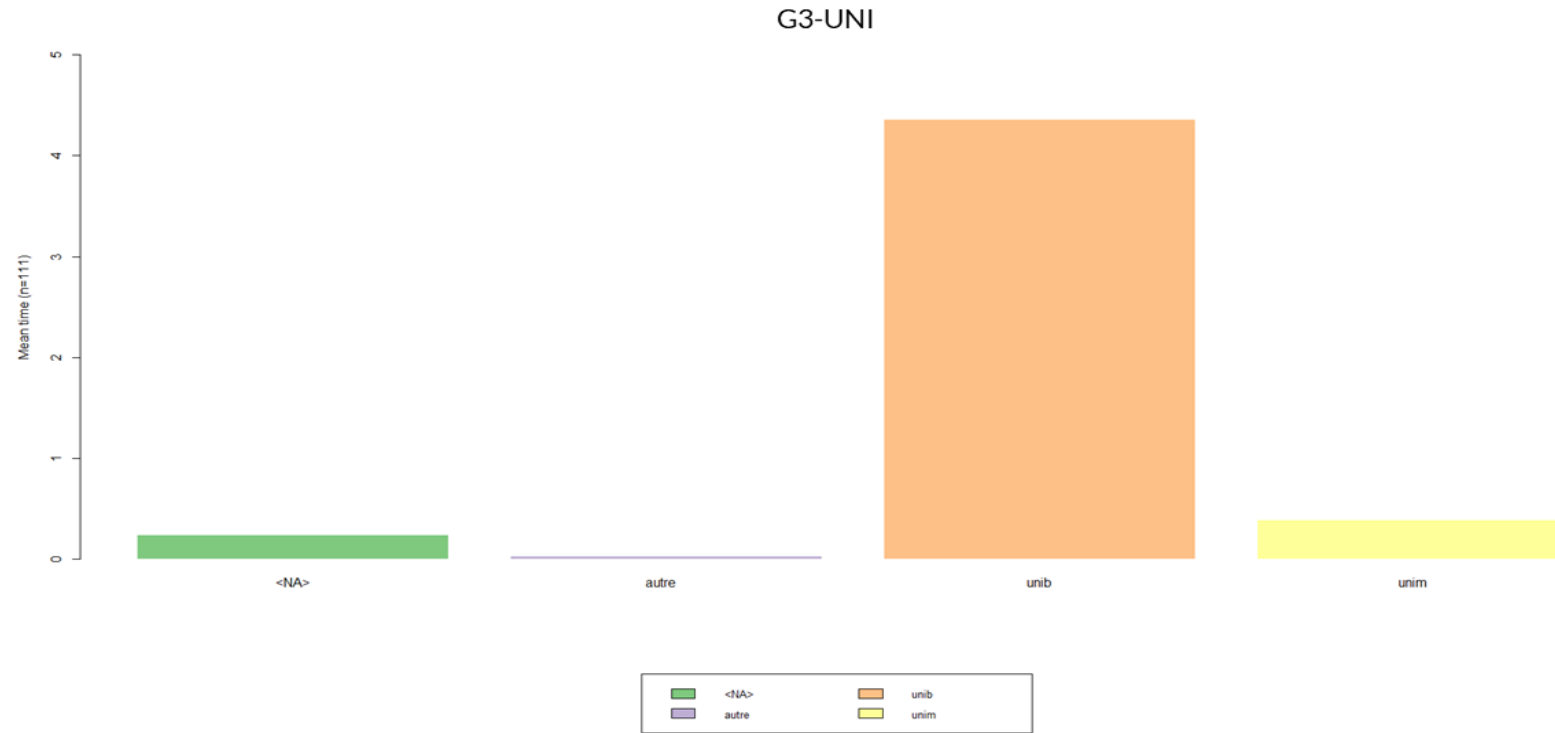


Figure 24. Temps moyen passé dans chaque état (unité de mesure = année)



8. LE TYPE INTERRUPTION DE FORMATION (N=62)

Figure 25. Séquences individuelles

G4-NA

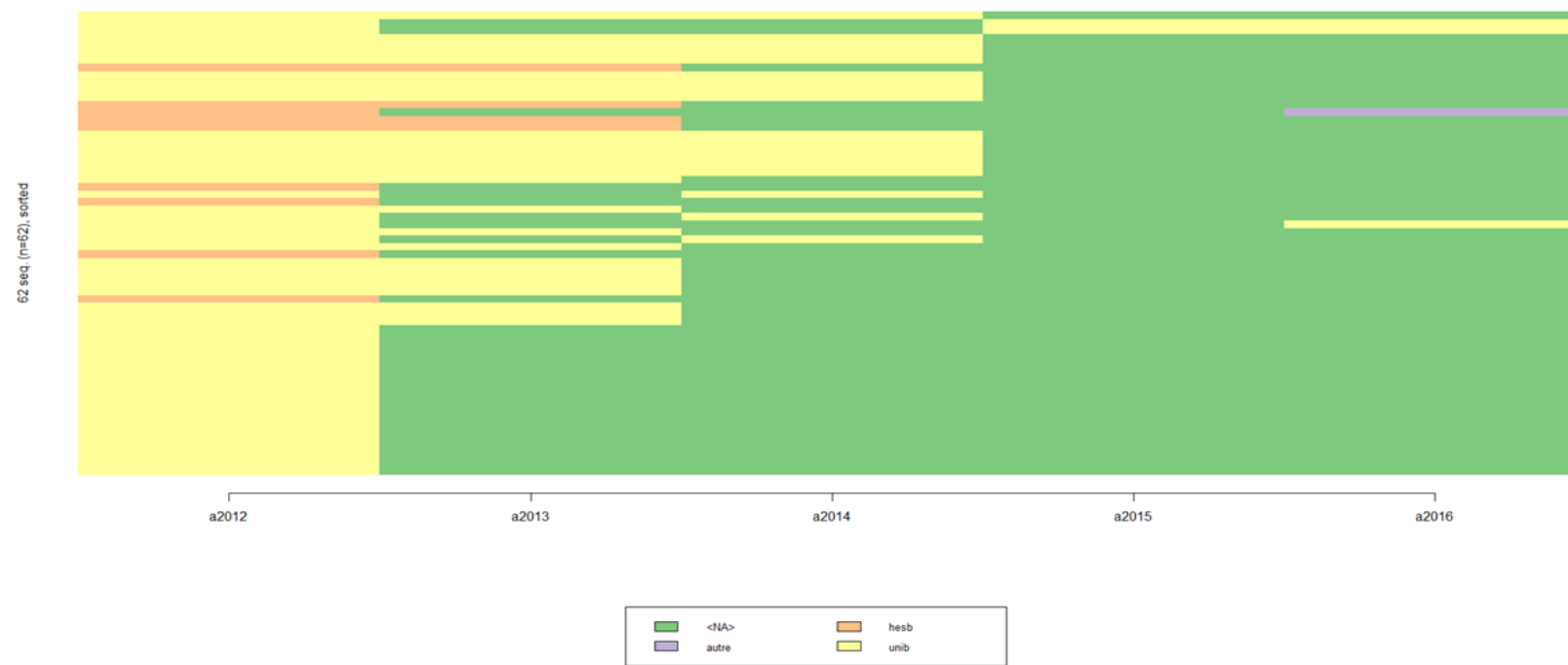


Figure 26. Séquences transversales

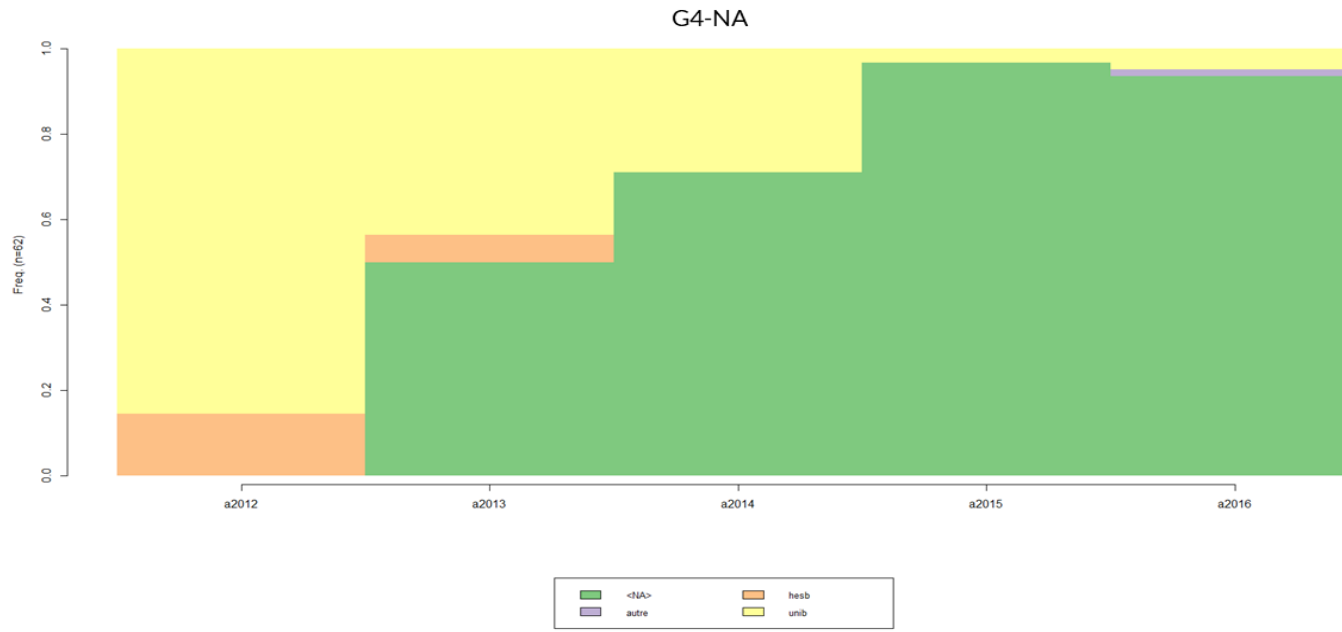


Figure 27. Mesure de diversité : entropie transversale

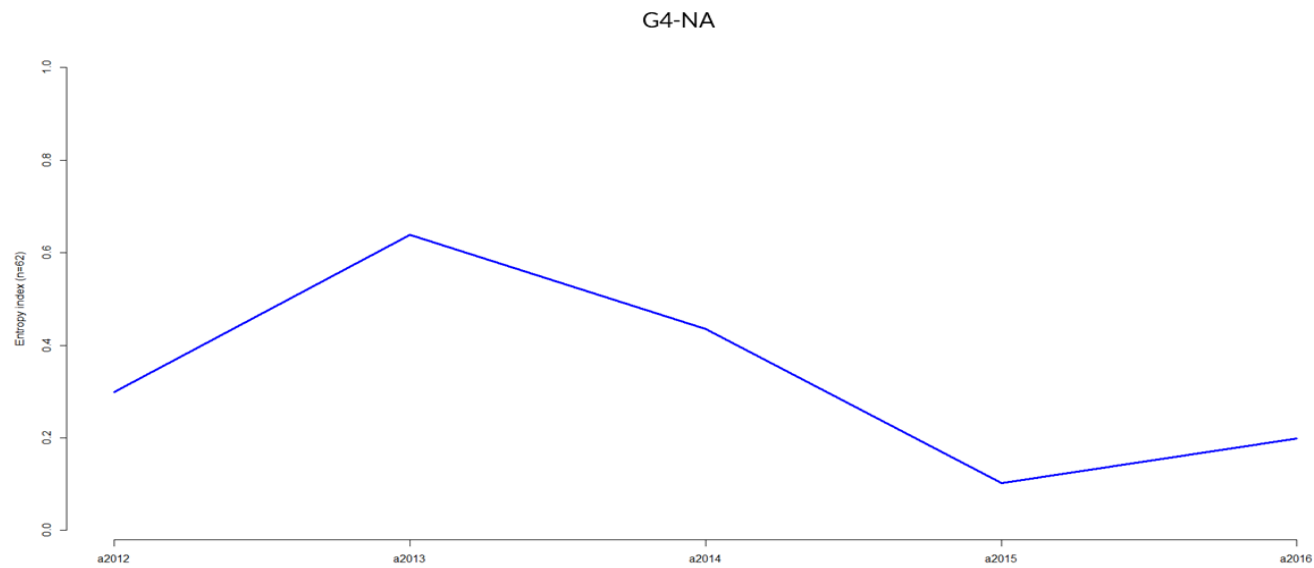
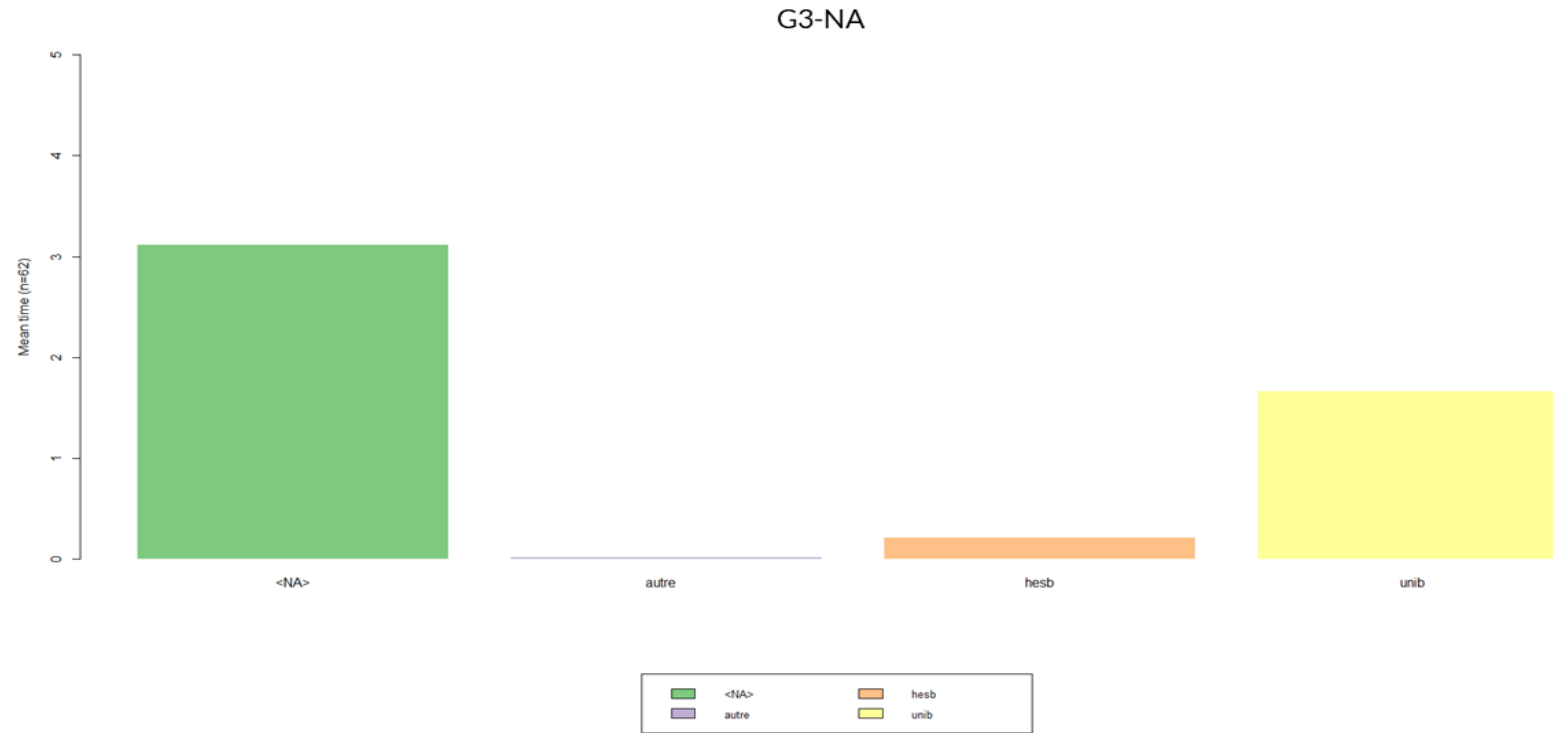


Figure 28. Temps moyen passé dans chaque état (unité de mesure = année)



9. BIBLIOGRAPHIE

Blanchard, Philippe, Bühlmann, Felix et Jacques-Antoine Gauthier. 2014. *Advances in sequence analysis: Theory, methods, applications* (First., p. 304). Amsterdam : Springer.

Delcher, Arthur L., Simon Kasif, Robert D. Fleischmann, Jeremy Peterson, Owen White et Steven L. Salzberg. 1999. Alignment of whole genomes. *Nucleic Acid Research* 27(11) : 2369-2376.

Gabardinho, Alexis, Gilbert Ritschard, Matthias Studer et Nicolas Müller. 2009. *Mining sequence data in R with TraMineR: A user's guide*. Technical report. Department of Econometrics and Laboratory of Demography, University of Geneva.

Robette, Nicolas. 2012. L'analyse de séquences : une introduction avec le logiciel R et le package TraMineR. *QUANTI/sciences sociales* [En ligne], consulté le 2 novembre 2017. URL : [ttp://quanti.hypotheses.org/686](http://quanti.hypotheses.org/686)

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technological Journal* 27(4) : 623-656.

Waterman, Michael S. 1995. *Introduction to Computational Biology. Maps, Sequences and Genomes*. London : Chapman et Hall.

