

# Anàlisi computacional de la microbiota intestinal i el comportament motor del nematode *Caenorhabditis elegans*

**Eric Alegret Olivé**

Màster en Bioinformàtica i Bioestadística

Àrea 4

**Teresa Torres Moral**

**(Antoni Pérez Navarro)**

Data d'entrega: 02 de juny de 2022



Esta obra está sujeta a una licencia de Reconocimiento-  
NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	Anàlisi computacional de la microbiota intestinal i el comportament motor del nematode <i>Caenorhabditis elegans</i> .
<b>Nom de l'autor:</b>	Eric Alegret Olivé
<b>Nom del consultor/a:</b>	Teresa Torres Moral
<b>Nom del PRA:</b>	Antoni Pérez Navarro
<b>Data d'entrega (mm/aaaa):</b>	06/2022
<b>Titulació:</b>	Màster en Bioinformàtica i Bioestadística
<b>Àrea del Treball Final:</b>	Àrea 4
<b>Idioma del treball:</b>	Català
<b>Número de crèdits:</b>	15
<b>Paraules clau</b>	<i>Caenorhabditis elegans</i> , microbiota entèrica, ecologia del moviment
<b>Resum del treball (màxim 250 paraules):</b> <i>Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball</i>	
<p>El nematode <i>Caenorhabditis elegans</i> és considerat actualment un model efectiu per a la recerca experimental i biomèdica, ja que posseeix un desenvolupament ràpid, temps de generació curts i una elevada homologia genètica i molecular amb humans. No obstant això, alguns aspectes de la seva biologia fonamental encara no s'han elucidat de forma precisa, com són el seus patrons de moviment i la composició de la seva comunitat microbiana intestinal. Així doncs, en aquest projecte s'ha enfocat l'estudi de <i>C. elegans</i> des d'aquestes dues perspectives clarament diferenciades. D'una banda, s'ha elaborat un codi reproduïble útil que combina llenguatge GNU-Bash i R per tal de dur a terme un meta-anàlisi amb vuit <i>datasets</i> de seqüències disponibles públicament, i descriure així la microbiota entèrica d'alguns membres del gènere <i>Caenorhabditis</i>. D'altra banda, s'ha adaptat i optimitzat un <i>pipeline</i> bioinformàtic d'R apropiat per a la caracterització dels patrons locomotors de les soques N2 i CB4856 de <i>C. elegans</i>. Els resultats obtinguts de la primera part metodològica han permès contrastar la comunitat intestinal característica de diferents espècies de nematodes, mostrant una clara influència de l'entorn de creixement dels cucs en la composició de la seva microbiota. Per un altre costat, el comportament motor de <i>C. elegans</i> sembla establir-se segons una distribució de probabilitat mixta que recull els dos principals tipus de moviment: el serpenteig rectilini i les reorientacions. La possibilitat de caracteritzar aquests dos aspectes rellevants en la biologia del nematode suposen la base de futurs estudis en els quals podria inferir-se una possible interconnexió entre la microbiota intestinal i el moviment de <i>C. elegans</i>.</p>	

**Abstract (in English, 250 words or less):**

The nematode *Caenorhabditis elegans* is currently considered an effective model for experimental and biomedical research since it has fast development, short generation times and a high genetic and molecular homology with humans. However, some aspects of its fundamental biology, as its movement patterns and the composition of its intestinal microbiota, have not been clearly elucidated yet. Therefore, this project aims to study *C. elegans* from these two clearly differentiated perspectives. On the one hand, a useful reproducible code that combines GNU-Bash and R languages has been developed to carry out a meta-analysis with eight publicly available sequence datasets, in order to describe the enteric microbiota of some members of the *Caenorhabditis* genus. On the other hand, an appropriate R bioinformatic pipeline has been adapted and optimized for the characterization of the locomotor patterns of *C. elegans* N2 and CB4856 strains. The results obtained from the first methodological part have made it possible to contrast the characteristic intestinal community of different nematode species, showing a clear influence of the growth environment of worms on the composition of their microbiota. On the second one, the motor behaviour of *C. elegans* seems to be established according to a mixed probability distribution that includes the two main types of motion: rectilinear crawling and reorientation. The possibility of characterizing these two relevant aspects in the biology of the nematode is the basis for future studies in which a possible interconnection between the intestinal microbiota and the movement of *C. elegans* could be inferred.

# Índex

1	Resum .....	1
2	Introducció .....	1
2.1	Context i justificació del treball .....	1
2.2	Objectius .....	4
2.3	Enfoc i mètode a seguir .....	5
2.4	Planificació .....	5
2.5	Breu sumari de contribucions i productes obtinguts .....	8
2.6	Breu descripció dels capítols de la memòria .....	8
3	Estat de l'art .....	8
3.1	La microbiota intestinal de <i>C. elegans</i> .....	8
3.2	Mètodes d'estudi del comportament motor dels nematodes .....	9
4	Metodologia .....	10
4.1	Caracterització de la microbiota intestinal del nematode <i>C. elegans</i> .....	10
4.1.1	Mètodes i procediments del meta-anàlisi .....	11
A.	Descàrrega i descripció inicial dels <i>datasets</i> de seqüències .....	11
B.	Pre-processat inicial dels <i>datasets</i> : <i>merging</i> .....	13
C.	Pre-processat inicial dels <i>datasets</i> : anàlisi de qualitat .....	14
D.	Pre-processat inicial dels <i>datasets</i> : filtratge de seqüències .....	16
E.	Pre-processat conjunt dels <i>datasets</i> : identificació de seqüències úniques .....	17
F.	Pre-processat conjunt dels <i>datasets</i> : generació d'OTUs al 97% d'identitat .....	17
G.	Pre-processat conjunt dels <i>datasets</i> : classificació taxonòmica dels OTUs .....	19
H.	Anàlisi descriptiva i estadística de les dades pre-processades .....	20
4.2	Estudi i modelització computacional dels patrons motors del nematode <i>C. elegans</i> .....	21
4.2.1	Mètodes i procediments seguits .....	21
A.	Fase experimental I: preparació de reactius i preservació de <i>stocks</i> de nematodes .....	21
B.	Fase experimental II: pre-cultiu d' <i>E. coli</i> i preparació de plaques NGM .....	21
C.	Fase experimental III: descongelació de vials i sincronització de nematodes ..	22
D.	Generació de trajectòries individuals d'alta resolució .....	22
E.	Simplificació, optimització i implementació d'un <i>pipeline</i> bioinformàtic per a la caracterització de trajectòries .....	23
5	Resultats .....	27
5.1	Descripció del perfil microbià intestinal del gènere <i>Caenorhabditis</i> .....	27
5.2	Caracterització dels patrons de locomoció de <i>C. elegans</i> .....	33
6	Discussió .....	38
6.1	Variació del microbioma entèric del gènere <i>Caenorhabditis</i> .....	38
6.2	Contrast de motilitat entre diverses soques de <i>C. elegans</i> .....	39

7	Conclusions .....	40
7.1	Conclusions .....	40
7.2	Línies de futur .....	41
7.3	Seguiment de la planificació .....	41
8	Glossari.....	42
	Bibliografia .....	43
	Annexes .....	48
	ANNEX I: Gràfics de qualitat dels <i>datasets</i> .....	48
	ANNEX II: Gràfics de locomoció de la resta d'individus.....	54
	ANNEX III: Codi generat .....	65

## Llista de figures

Figura 1. Diagrama de Gantt amb la temporització de les diferents tasques del present Treball de Fi de Màster (indicades a la columna de la part esquerra de la imatge). .....	7
Figura 2. Representació gràfica generada amb R dels estadístics de qualitat associats a les lectures d'un metagenoma (només es presenten a tall d'exemple per al dataset Dirksen_native). A) Distribució de la longitud dels reads de seqüenciació (en pb); B) Distribució dels índexs de qualitat de les lectures obtingudes; C ) Distribució dels índexs de qualitat en funció de la longitud dels reads de seqüenciació (en pb); D) Distribució de l'expected error promig en funció de la longitud de seqüència (en pb); E) Distribució percentual del nº de seqüències preservades després d'implementar un filtratge en base a l'expected error i la longitud dels reads (en pb). .....	15
Figura 3. Gràfic NMDS de les mostres de microbioma identificades segons la condició de creixement de l'hoste de procedència (Growing Condition). .....	28
Figura 4. Gràfic NMDS de les mostres de microbioma identificades segons la condició de creixement de l'hoste de procedència (Growing Condition), després de resumir les categories i eliminar les mostres de microbioma de substrat. ....	28
Figura 5. Gràfic NMDS de les mostres de microbioma identificades segons l'hoste de procedència i la seva condició de creixement, després de resumir les categories i eliminar les mostres de microbioma del substrat. ....	29
Figura 6. A) Boxplot múltiple per estudiar la dispersió de les mostres i la distància d'aquestes al centroide de l'el·lipse que els conté (eix d'abscisses), en base a les diferents combinacions d'espècie de metazou i condició de proliferació (eix d'ordenades). B) Càlcul resultant del test de Tukey per a la comparació de distàncies mitjanes al centroide entre les diverses combinacions de grups valorades. ....	30
Figura 7. Representació de la comunitat intestinal de <i>C. elegans</i> diferenciant les diferents mostres segons l'entorn de creixement de l'individu de procedència, indicant les classificacions taxonòmiques dels OTUs en base al color descrit a la llegenda. A més, es mostren enumerats els tàxons amb una elevada abundància relativa en una o més condicions de creixement. ....	31
Figura 8. Representació de la comunitat intestinal de <i>C. elegans</i> , <i>C. remanei</i> i <i>C. briggsae</i> aïllats de l'entorn natural, indicant les classificacions taxonòmiques dels OTUs en base al color descrit a la llegenda. A més, es mostren enumerats els tàxons amb una elevada abundància relativa en una o més espècies. ....	31
Figura 9. Diagrama de dispersió que il·lustra l'abundància relativa mitjana dels OTUs (eix d'ordenades) representats en base al nº de mostres avaluades (eix d'abscisses) de <i>C. elegans</i> . S'indiquen els gèneres bacterians presents en almenys el 70% de les mostres amb una abundància major al 0.5%, indicant la família corresponent en base al color. ....	32
Figura 10. Diagrama de dispersió que il·lustra l'abundància relativa mitjana dels OTUs (eix d'ordenades) representats en base al nº de mostres avaluades (eix d'abscisses) de <i>C. remanei</i> . S'indiquen els gèneres bacterians presents en almenys el 70% de les mostres amb una abundància major al 0.5%, indicant la família corresponent en base al color. ....	33
Figura 11. Representació de la trajectòria dels individus 201405101451 (soca N2) i 201406041832 (soca CB4856) de <i>C. elegans</i> en base a les coordenades cartesianes X/Y mesurades en mil·límetres. ....	33
Figura 12. Distribucions de densitat de les velocitats lineals seguides per l'individu 201405101451 de la soca N2 de <i>C. elegans</i> . A) Representació de la distribució mixta (en vermell) i individual (en blau) dels segments de trajectòria assignats a la distribució probabilística gamma. B) Representació de la distribució mixta (en vermell) i individual (en groc) dels segments de	

trajectòria assignats a la distribució probabilística gaussiana. C) Il·lustració conjunta de la distribució combinada diferenciant per colors cadascuna de les parts elementals que la constitueixen. S'indiquen addicionalment els paràmetres que defineixen la distribució gamma (shape i rate) i normal (mean i sd). .....	34
Figura 13. Distribucions de densitat de les velocitats lineals seguides per l'individu 201406041832 de la soca CB4856 de <i>C. elegans</i> . A) Representació de la distribució mixta (en vermell) i individual (en blau) dels segments de trajectòria assignats a la distribució probabilística gamma. B) Representació de la distribució mixta (en vermell) i individual (en groc) dels segments de trajectòria assignats a la distribució probabilística gaussiana. C) Il·lustració conjunta de la distribució combinada diferenciant per colors cadascuna de les parts elementals que la constitueixen. S'indiquen addicionalment els paràmetres que defineixen la distribució gamma (shape i rate) i normal (mean i sd). .....	35
Figura 14. A) Diagrama d'etograma que descriu els estats de reorientació (en blau) i crawling (en groc) de l'individu indicat (soca N2), en base a la columna state de la taula de resultats (dreta). B) Gràfic de barres que indica la velocitat associada a cadascun dels estats motors de la trajectòria.....	35
Figura 15. A) Diagrama d'etograma que descriu els estats de reorientació (en blau) i crawling (en groc) de l'individu indicat (soca CB4856), en base a la columna state de la taula de resultats (dreta). B) Gràfic de barres que indica la velocitat associada a cadascun dels estats motors de la trajectòria.....	36
Figura 16. Distribucions de densitat (eix d'ordenades) de les velocitats angulars en rad/s (eix d'abscisses) de l'individu indicat (soca N2), distingint els valors designats al moviment de crawling (ocre) i de reorientació o looping (blau). A la dreta es mostra el model computat de velocitat angular que identifica els dos tipus de patró locomotor. ....	36
Figura 17. Distribucions de densitat (eix d'ordenades) de les velocitats angulars en rad/s (eix d'abscisses) de l'individu indicat (soca CB456), distingint els valors designats al moviment de crawling (ocre) i de reorientació o looping (blau). A la dreta es mostra el model computat de velocitat angular que identifica els dos tipus de patró locomotor. ....	37



## **Lista de taules**

Taula 1. Resum dels paràmetres de les equacions de Langevin calculats per als individus emprats com a exemple de les dues soques d'estudi. ....	37
---	----

# 1 Resum

El nematode *Caenorhabditis elegans* és considerat actualment un model efectiu per a la recerca experimental i biomèdica, ja que posseeix un desenvolupament ràpid, temps de generació curts i una elevada homologia genètica i molecular amb humans. No obstant això, alguns aspectes de la seva biologia fonamental encara no s'han elucidat de forma precisa, com són el seus patrons de moviment i la composició de la seva comunitat microbiana intestinal. Així doncs, en aquest projecte s'ha enfocat l'estudi de *C. elegans* des d'aquestes dues perspectives clarament diferenciades. D'una banda, s'ha elaborat un codi reproduïble útil que combina llenguatge GNU-Bash i R per tal de dur a terme un meta-anàlisi amb vuit *datasets* de seqüències disponibles públicament, i descriure així la microbiota entèrica d'alguns membres del gènere *Caenorhabditis*. D'altra banda, s'ha adaptat i optimitzat un *pipeline* bioinformàtic d'R apropiat per a la caracterització dels patrons locomotors de les soques N2 i CB4856 de *C. elegans*. Els resultats obtinguts de la primera part metodològica han permès contrastar la comunitat intestinal característica de diferents espècies de nematodes, mostrant una clara influència de l'entorn de creixement dels cucs en la composició de la seva microbiota. Per un altre costat, el comportament motor de *C. elegans* sembla establir-se segons una distribució de probabilitat mixta que recull els dos principals tipus de moviment: el serpenteig rectilini i les reorientacions. La possibilitat de caracteritzar aquests dos aspectes rellevants en la biologia del nematode suposen la base de futurs estudis en els quals podria inferir-se una possible interconnexió entre la microbiota intestinal i el moviment de *C. elegans*.

## 2 Introducció

### 2.1 Context i justificació del treball

El projecte d'enguany es contextualitza al voltant del metazou *Caenorhabditis elegans*, un petit nematode vermiforme de vida lliure, no parasitari i no patogènic, propi d'ambients terrestres de climes temperats <sup>[1,2]</sup>. Avui dia, *C. elegans* és considerat un organisme model efectiu per a la recerca experimental i biomèdica, especialment en àrees tals com la genètica, la farmacologia, la neurobiologia i la gerontologia <sup>[3,4]</sup>. Fou introduït com a animal d'experimentació l'any 1963 per Sydney Brenner amb l'objectiu de caracteritzar processos de la biologia eucariota des d'un punt de vista molecular, i des d'aleshores s'ha convertit en el focus d'experimentació i nous descobriments de molts laboratoris arreu del món <sup>[5,6,7]</sup>. No obstant això, alguns aspectes de la seva biologia fonamental encara no s'han elucidat de forma precisa, com són el seus patrons de moviment i la composició de la seva comunitat microbiana intestinal.

El nematode en qüestió posseeix un desenvolupament ràpid, temps de generació curts i una elevada homologia genètica i molecular amb humans <sup>[3,5,8,9]</sup>. A més, prolifera contínuament i de forma diària, és senzill de mantenir i manipular al laboratori i no està subjecte a consideracions ètiques <sup>[9]</sup>. Tot i que és anatòmicament més simple, estudis recents semblen haver demostrat la conservació a *C. elegans* d'un elevat número de gens i rutes patològiques pròpies d'humans i altres mamífers <sup>[3,6,9]</sup>. És per tot això que *C. elegans* s'empra per estudiar multitud de processos cel·lulars, incloent la regulació gènica, el metabolisme, l'apoptosi, el cicle cel·lular o els sistemes de senyalització <sup>[3,9]</sup>. Tot això evidencia el seu enorme potencial per a la realització de *high-throughput screenings* (cribratges d'alt rendiment basats en l'ús d'equipament automatitzat útil pel testeig simultani de l'activitat biològica d'un elevat nº de mostres, extractes o compostos) <sup>[3,9,10]</sup>. Així doncs, en aquest projecte es pretén enfocar l'estudi de *C. elegans* des de dues perspectives clarament diferenciades: l'anàlisi bioinformàtic de la seva microbiota intestinal i la caracterització computacional del seu comportament motor.

### **Breu caracterització del model *Caenorhabditis elegans***

*C. elegans* pertany al grup dels nematodes, descrits col·loquialment com a cucs rodons o cilíndrics. Es classifica taxonòmicament dins la família *Rhabditidae*, inclosa en el filum *Nematoda*, el superfilum *Ecdysozoa* i els clades sense valor taxonòmic *Protostomia* i *Bilateria* <sup>[11,12]</sup>. Així doncs, *C. elegans* conserva les principals sinapomorfies d'aquests nivells jeràrquics de classificació biològica <sup>[11,12]</sup>:

- És un organisme amb simetria bilateral i, per tant, presenta un pla sagital de simetria corporal que permet definir un eix antero-posterior fàcilment identificable <sup>[11,12]</sup>.
- Té un desenvolupament embrionari triblàstic basat en l'aparició de tres capes cel·lulars: l'ectoderma, que donarà lloc a l'epidermis i el sistema nerviós; el mesoderma, que originarà la musculatura, les gònades i el cor, i l'endoderma, a partir del qual es formarà el tub digestiu i el sistema excretor <sup>[11,12]</sup>.
- El zigot pateix segmentació en espiral durant el seu temps de proliferació <sup>[11,12]</sup>.
- La boca de l'individu, denominada estomodeu, es forma a partir del blastòpor, una invaginació cel·lular produïda durant la formació de l'embrió, que en alguns animals pot originar un tub digestiu cec <sup>[11,12]</sup>.
- Presenta esquizocèlia, és a dir, proliferació cel·lular entre l'ecto- i l'endoderma que pot arribar a desenvolupar el mesoderma, i conseqüentment el corresponent celoma que propiciarà la formació d'epiteli intern entre les vísceres <sup>[11,12]</sup>.
- El seu cos no conté cilis i es troba recobert d'una cutícula de tres capes que canvien durant el procés d'ècdisi o muda <sup>[11,12]</sup>.

Tots els individus d'aquesta espècie són blastocelomats; això vol dir que la cavitat tubular del seu cos no és més que un blastocel (terme emprat per descriure la primera cavitat embrionària formada en el zigot) que persisteix en l'organisme adult i ocupa la major part de l'espai intern, ple de líquid i amb els òrgans d'un sistema digestiu complet <sup>[11,12]</sup>. A més, la seva estructura corporal és translúcida i de secció circular insegmentada, amb només musculatura longitudinal que li permet seguir trajectòries i moviments ondulatoris <sup>[11,12]</sup>. Cal afegir que es tracta d'un metazou amb poca cefalització, ja que disposa d'un sistema nerviós simple constituït per un gangli cerebroide poc complex i un baix número de neurones (al voltant de 300) <sup>[11,12,13,14]</sup>. Tampoc té sistema circulatori ni respiratori, doncs intercanvia gasos per difusió a través de la cutícula <sup>[11,12]</sup>.

El nematode *C. elegans* presenta una àmplia distribució global i regional, i resulta ser un habitant típic tant de sòls humits amb un nivell moderat d'oxigen com de matèria vegetal en descomposició <sup>[2,13,14]</sup>. Té uns hàbits alimentaris principalment bacteriòvors, ja que consumeix per ingesta directa organismes procarïotes en suspensió o propis del detritus <sup>[1]</sup>. Addicionalment, al voltant de les estructures bucals aglutina diversos òrgans sensorials de caràcter mecanoreceptor, quimiorceptor o fotoreceptor, els quals l'ajuden a guiar-se en l'entorn a l'hora de cercar aliment, evitar toxines o depredadors, o bé aparellar-se <sup>[2,7,11,12]</sup>.

A nivell genètic, els individus de *C. elegans* tenen un cariotip format per cinc autosomes (I a V) i un cromosoma sexual (X) <sup>[2,7]</sup>. Conseqüentment, l'espècie adopta dos possibles sexes: mascle (XO) o hermafrodita (XX) (la gran majoria, >80% de la població), impossibilitant l'aparició de femelles de forma natural <sup>[1,2]</sup>. La descendència generada per auto-fecundació dels individus hermafrodites és considerable (uns 300 ous), malgrat que l'aparellament amb mascles a través de les espícules propicia números encara més elevats <sup>[2]</sup>. Per tant, dels cucs hermafrodites s'originen poblacions genèticament idèntiques formades per individus clònics <sup>[2]</sup>. D'altra banda, el genoma de *C. elegans* té una mida aproximada de 100 Mb i conté al voltant de 20.000 gens, alguns dels quals estan agrupats en operons tot i ser una particularitat pròpia del món procarïota <sup>[2]</sup>.

El cicle vital de *C. elegans* es completa en aproximadament 3 dies a una temperatura d'incubació de 20°C [2,7,14]. Durant aquest temps, els individus segueixen un creixement discontinu que passa per 4 estadis larvaris (L1 a L4) fins assolir l'etapa adulta capaç de pondre ous, els quals romanen estables com a màxim 3 setmanes [2,7,14]. Cal destacar que els nematodes adults poden arribar a mesurar 1 mm de longitud i 80 µm de diàmetre [2,7].

El gènere *Caenorhabditis* engloba més de 25 espècies diferents, totes elles amb nombrosos trets morfològics i fisiològics compartits que dificulten la seva identificació fenotípica [2]. Tot i així, aquestes són clarament diferenciables a nivell genotípic [2].

### **La microbiota intestinal humana i el *gut-brain axis*. Aproximacions plausibles amb el model *C. elegans***

En la darrera dècada, disciplines com la Microbiologia i la Neurociència han convergit en la caracterització de les interaccions dinàmiques que es donen entre animals i les seves comunitats bacterianes residents, les quals contribueixen a l'establiment i funcionalització dels sistemes neurològics [15,16]. La microbiota intestinal, o *gut microbiota*, es defineix com el conjunt d'espècies microbianes (comensals, simbiotes i patògeniques) que ocupen els diferents nínxols ecològics de les superfícies mucoses del tracte gastrointestinal [17,18,19,20,21]. Es tracta d'una entitat dinàmica formada principalment per bacteris, però també inclou virus, arqueus, fongs, protozous i paràsits [17,18,19,20,21].

L'ecosistema intestinal humà presenta una àmplia diversitat i abundància bacteriana al llarg del tracte, que alberga més de 1000 espècies distribuïdes en més de 50 fílums [18,22]. D'aquests últims, els predominants són *Bacteroidetes* i *Firmicutes*, seguits de *Proteobacteria*, *Fusobacteria*, *Tenericutes*, *Actinobacteria*, *Verrucomicrobia* i *Cyanobacteria* [21]. Tanmateix, la proporció dels diferents fílums microbianos difereix entre individus en funció de factors intrínsecs [22] (genètica i epigenètica de l'hoste [16,23,24], ritme circadià [16]) i extrínsecs (modalitat de part [16,24], dieta [16,24], ambient [16,24], rutina d'activitat física [16,24], tractaments mèdics [16,24], estrès [16,24], edat [24,25]).

La microbiota intestinal duu a terme funcions importants pel manteniment fisiològic de l'organisme, com colonitzar les superfícies del tracte digestiu i consolidar la integritat de la barrera intestinal; protegir front patògens a través de la secreció de productes antimicrobians; madurar la resposta immunològica; degradar xenobiòtics i intervenir en la digestió i el metabolisme; controlar la proliferació i diferenciació de cèl·lules epitelials, i afavorir la síntesi de molècules essencials (com hormones i vitamines) [16,18,22]. Canvis anormals en la composició de la *gut microbiota* (l'anomenada disbiosi) poden conduir a la patofisiologia de dolències com asma, diabetis, insuficiència renal o malalties neurològiques [19,22].

*C. elegans* s'està avaluant com a model factible per a la caracterització del perfil microbià que colonitza l'intestí [8,26]. En els darrers anys s'ha posat de manifest el rol cabdal de la microbiota intestinal en el desenvolupament cognitiu, el comportament i l'homeòstasi, generalment a través de la producció i modulació de factors metabòlics, immunològics i neuroquímics [16,27]. Aquesta nova perspectiva ha permès correlacionar la composició i funció de la microbiota entèrica amb els mecanismes subjacents de molts desordres neuropsiquiàtrics i neurodegeneratius, provocant un canvi de paradigma quant a l'impacte que exerceixen els microorganismes de l'intestí en la salut i en la malaltia [15,16,27]. El terme *gut-brain axis* (GBA) designa la complexa xarxa d'interconnexions que permeten la comunicació bidireccional entre el tracte gastrointestinal i el cervell [15]. S'ha demostrat, però, que la microbiota intestinal pot modular la senyalització d'aquest eix, definit ara com a *microbiota-gut-brain axis* [15]. Considerant els avantatges de treballar amb *C. elegans* descrits anteriorment, cada cop més es contempla la possibilitat d'utilitzar el nematode com a model per fer aproximacions plausibles d'aquest eix [26,28]. Interpretar els canvis en la microbiota intestinal d'aquest

organisme permetria, idealment, extrapolar certes conclusions i aplicar-les a humans, la qual cosa suposaria un canvi revolucionari en la descripció de l'etiologia de moltes patologies <sup>[26,28]</sup>.

### **El comportament animal i l'ecologia del moviment**

Identificar i descriure els principals impulsors del comportament de cerca animal roman com a repte destacable dins la branca de l'ecologia del moviment <sup>[29]</sup>. Farratjar (o *foraging* en anglès) fa referència al conjunt de conductes centrades en l'exploració, la cerca i obtenció d'aliment o substrat originades com a resposta a l'ambient en què es troba l'animal, el qual és capaç de generar complexes estratègies exploratòries per assolir-lo <sup>[29,30]</sup>. Aquesta tàctica multifacètica implica una detecció prèvia de l'aliment per cerca i taxis, així com esdeveniments posteriors de persecució i/o manipulació de la presa <sup>[30,31]</sup>. La resposta animal, doncs, ve influenciada tant per les característiques de l'entorn (espacio-temporals, principalment) com per l'aprenentatge basat en experiències anteriors i l'estat en què es trobi l'individu <sup>[29,30,31]</sup>.

Malgrat això, els estímuls ambientals no són sempre els únics factors causals de les reaccions motores de l'animal, ja que la informació procedent del medi pot resultar en alguns casos difusa, de qualitat variable, intermitent (estímuls percebuts en intervals de temps) o directament inexistent <sup>[30,32]</sup>.

L'entorn ha de ser explorat adequadament per poder explotar de forma eficient els recursos que hi hagi disponibles <sup>[30]</sup>. Ara bé, si les senyals mediambientals són escasses, els animals es reorienten per trobar un equilibri entre la cerca intensiva (en l'entorn més proper) i extensiva (en les zones més llunyanes respecte la seva posició) <sup>[29,33]</sup>. Aquesta dinàmica de comportament es coneix com *random search theory*, i ve determinada per tres grans factors: 1) la rapidesa de l'animal, 2) els patrons d'orientació i locomoció, i 3) la percepció de l'individu <sup>[29,33]</sup>. Tots plegats optimitzen la cerca d'aliment local i distant quan es desconeix la seva distribució espacial <sup>[34,35]</sup>.

De tota manera, és realment complicat poder estudiar trajectòries animals en espais naturals oberts amb la finalitat de desenvolupar models estocàstics dels seus patrons motors <sup>[30]</sup>. En aquestes condicions es fa difícil enregistrar els *inputs* sensorials que els organismes reben, així com l'estatus perceptiu que l'animal pugui tenir <sup>[30,36]</sup>. La qualitat del mostreig, el número de mesures i l'anàlisi estadística a aplicar sobre els conjunts de dades de moviment animal esdevenen factors crítics a l'hora de desenvolupar estudis d'ecologia del comportament <sup>[30,36]</sup>. Aquestes dades, idealment, s'haurien de recollir contínuament i amb una elevada freqüència durant grans períodes de temps, la qual cosa permetria generar resultats estadísticament significatius <sup>[30,36]</sup>.

*C. elegans*, donades les seves qualitats com a organisme model, resulta idoni per generar experiments d'aquest caire en espais controlats al laboratori i definir models heurístics que expliquin intricades accions motores, algunes de les quals podrien, fins i tot, venir influenciades en menor o major grau per la microbiota intestinal. Estudiar la dinàmica motora difusòria que pren el nematode en experiments manipulatius i la posterior anàlisi computacional de les dades obtingudes permetrà, sens dubte, millorar la concepció dels patrons de moviment animal <sup>[29,37]</sup>.

## **2.2 Objectius**

### **Objectius generals**

- Descriure el perfil de la comunitat microbiana intestinal del nematode *C. elegans* i el d'altres soques i/o espècies properes.
- Caracteritzar el comportament motor de *C. elegans* i la seva variabilitat entre soques i entre individus d'una mateixa soca.

### **Objectius específics**

- Generar un *pipeline* bioinformàtic que permeti determinar la composició de la microbiota intestinal de diferents nematodes, partint de les dades presents a la literatura i bases de dades públiques.
- Identificar el *core microbiome* intestinal natiu associat a algunes espècies del gènere *Caenorhabditis*.
- Reportar la dispersió de la microbiota entèrica natural de les espècies avaluades i, idealment, estudiar possibles diferències segons les condicions de creixement del cuc.
- Descriure i ajustar, a partir de trajectòries d'alta resolució (20 Hz) i mitjançant eines computacionals, els principals paràmetres que determinen els patrons motors de *C. elegans* segons un model d'equacions diferencials estocàstiques, anomenat model de Langevin.
- Idealment, contrastar la motilitat entre soques de *C. elegans* per veure com difereix el patró de moviment.

### 2.3 Enfoc i mètode a seguir

La seqüenciació d'alt rendiment d'amplicons del gen *16S* rDNA és una estratègia àmpliament emprada per a l'estudi de comunitats microbianes <sup>[38]</sup>. No obstant això, la identificació dels diferents grups microbians pot resultar més o menys precisa en funció del *pipeline* bioinformàtic aplicat <sup>[38]</sup>. A l'actualitat, existeixen nombroses eines computacionals adequades per a l'anàlisi de dades de microbioma, cadascuna de les quals implementa diferents mètodes d'estimació i assignació taxonòmica <sup>[38,39]</sup>.

El primer pas consistirà en recopilar, mitjançant el motor de cerca PubMed de NCBI (*National Center for Biotechnology Information*), la informació disponible sobre la caracterització de la microbiota intestinal dels nematodes del gènere *Caenorhabditis*, fent especial èmfasi en *C. elegans* i els tàxons més propers (ja siguin espècies o soques). Això conduirà a l'obtenció de les dades crues provinents de la seqüenciació del gen *16S* de les comunitats intestinals d'aquests cucs.

A continuació, s'executarà un pre-processat de les dades amb diversos *scripts* de llenguatge GNU-Bash per tal de filtrar i classificar les seqüències en *Operational Taxonomic Units* (OTUs). Per fer-ho, s'utilitzaran algoritmes d'elevada eficiència computacional com USEARCH <sup>[39]</sup>, VSEARCH <sup>[40]</sup> i UPARSE <sup>[41]</sup>. Seguidament, i partint dels OTUs generats en el pas previ, es realitzarà l'anàlisi descriptiva de la microbiota entèrica i les proves bioestadístiques associades amb el programari lliure R, emprant alhora l'entorn de desenvolupament integrat (o IDE) RStudio per interactuar fàcilment amb les comandes <sup>[42,43]</sup>. Els diversos arxius de codi generats amb ambdós llenguatges constituiran, doncs, un *pipeline* de creació pròpia útil per l'estudi de la comunitat intestinal dels nematodes d'interès.

Pel que fa al bloc d'estudi del comportament motor dels nematodes, es començarà per obtenir fitxers de desplaçament dels cucs amb les corresponents coordenades espacials X/Y i anotacions temporals, mitjançant gravacions d'alta resolució del seu moviment. Aquestes trajectòries es caracteritzaran segons un model d'equacions diferencials estocàstiques basades en el model de Langevin <sup>[44]</sup>. Per aconseguir-ho, caldrà aplicar un algoritme computacional en llenguatge R per tal de distingir els diferents tipus de moviment (reptació, reorientació, piruetes, etc.) i calcular els paràmetres del model esmentat.

### 2.4 Planificació

#### Tasques

La següent planificació desglossa les diferents activitats que cal portar a terme durant el desenvolupament del present Treball de Fi de Màster. Entre parèntesis s'assigna la durada estimada de cada tasca.

1. Contextualització: cerca bibliogràfica per recopilar informació associada a la temàtica d'estudi.
  - 1.1. Revisió general del model *C. elegans*. (18h)

- 1.2. Revisió general sobre conceptes bàsics de la microbiota intestinal. (30h)
- 1.3. Revisió d'estudis experimentals centrats en el comportament motor animal. (24h)
2. Determinació dels objectius del treball i proposta de plantejament metodològic. (21h)

**BLOC MICROBIOLOGIA:** Caracterització de la microbiota entèrica del nematode *C. elegans*.

1. Recopilació d'articles sobre la comunitat microbiana intestinal del nematode d'estudi.
  - 1.1. Preparació de l'arxiu de *metadata* (individuals per a cada *dataset* i també global). (50h)
  - 1.2. Descàrrega i obtenció dels fitxers de seqüències de les bases de dades vinculades. (10h)
2. Diferenciació d'OTUs: Processat inicial del *dataset* de seqüències amb el clúster o supercomputadora, implementant *scripts* de llenguatge GNU-Bash.
  - 2.1. *Merging* de seqüències. (12h)
  - 2.2. Anàlisi de qualitat. (18h)
  - 2.3. Filtratge de les lectures. (15h)
  - 2.4. Identificació de seqüències úniques. (12h)
  - 2.5. Generació d'OTUs al 97% d'identitat. (20h)
  - 2.6. Classificació taxonòmica dels OTUs. (18h)
3. Anàlisi descriptiva de la microbiota entèrica, implementant un *script* en llenguatge R.
  - 3.1. Reducció del *dataset* i associació de les dades. (7h)
  - 3.2. Elaboració de gràfics descriptors de diversitat ecològica, taxonomia i abundància diferencial. (24h)
4. Proves bioestadístiques i anàlisi multivariant dels resultats assolits. (3h)

**BLOC ECOLOGIA DEL MOVIMENT:** Estudi i modelització dels patrons motors del nematode *C. elegans*.

1. Recopilació d'articles sobre la dinàmica motriu del nematode d'estudi. (20h)
2. Manipulació experimental i obtenció de trajectòries d'alta resolució a nivell individual en un sistema de seguiment de 90 min de duració cadascuna. (75h)
3. Implementació d'un algoritme computacional en llenguatge R per a la neteja de les trajectòries i la distinció dels diferents tipus de moviment. (80h)
4. Càlcul dels paràmetres motrius del model de Langevin. (24h)

**Calendari**

S'ha elaborat un diagrama de Gantt mitjançant l'eina GanttProject amb la finalitat d'esquematzar visualment les diferents tasques plantejades en l'apartat anterior, tot indicant la temporalització de cadascuna d'elles i també la de les pròpies PACs de l'assignatura (Fig. 1).

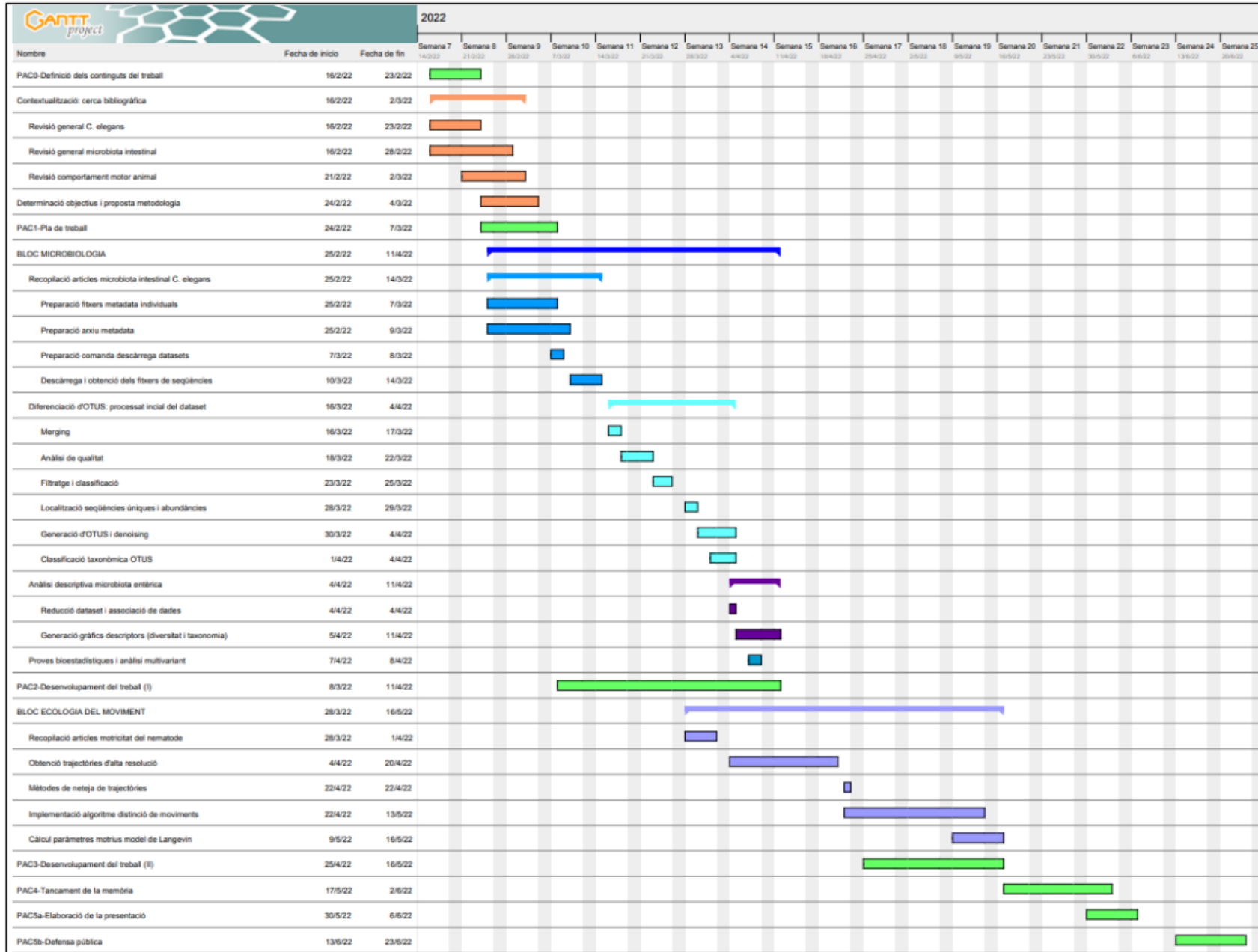


Figura 1. Diagrama de Gantt amb la temporització de les diferents tasques del present Treball de Fi de Màster (indicades a la columna de la part esquerra de la imatge).



**Fites**

El projecte plantejat presenta fites en ambdós blocs d'estudi. Assolir-los permetrà avançar en la consecució de les successives etapes de resultats pràctics.

**BLOC MICROBIOLOGIA**

- ✓ Descriure la qualitat de les seqüències del *raw data*.
- ✓ Obtenir la taula d'OTUs i el fitxer de seqüències representatives en el processat inicial amb GNU-Bash.
- ✓ Caracteritzar taxonòmicament amb R la microbiota intestinal dels nematodes i inferir diferències estadísticament contrastades quant a la composició microbiana de les diferents espècies o soques analitzades.

**BLOC ECOLOGIA DEL MOVIMENT**

- ✓ Obtenir els fitxers de trajectòries individuals (idealment 10 o més), a ser possible amb diverses soques de *C. elegans*.
- ✓ Diferenciar i segmentar els moviments de reptació dels moviments de reorientació (girs, voltes o piruetes) gràcies a l'aplicació de l'algoritme computacional.
- ✓ Estimar els paràmetres del model de Langevin.

**2.5 Breu sumari de contribucions i productes obtinguts**

L'assoliment dels objectius plantejats en aquest treball ha permès obtenir una memòria final on queda registrada la metodologia d'anàlisi bioinformàtica i els principals resultats obtinguts d'aquesta pel que fa a la caracterització del perfil de la microbiota entèrica del nematode d'estudi i la modelització de la seva locomoció. L'organització de les tasques i el subseqüent pla de treball defineixen l'abast i extensió del projecte. Així mateix, l'estudi ha finalitzat amb l'entrega d'una presentació en format virtual on s'exposen els aspectes més rellevants de la investigació.

Com a resultat final s'ha obtingut un *pipeline* de creació pròpia per a cadascun dels blocs d'estudi, un per a la caracterització de microbiotes (amb les corresponents classificacions taxonòmiques i els gràfics descriptors de diversitat) i un altre per a la modelització de trajectòries; pel que fa a aquest últim, també es generaran taules amb paràmetres modelitzats i algunes simulacions de moviment.

**2.6 Breu descripció dels capítols de la memòria**

Al capítol 3 de la present memòria de treball es descriuen alguns aspectes sobre l'estat de l'art actual dels dos àmbits d'estudi en els quals s'ha dividit el projecte, és a dir, la microbiota entèrica de *C. elegans* i l'anàlisi de la seva locomoció. A continuació, el capítol 4 detalla les diverses etapes procedimentals seguides per tal d'assolir els objectius plantejats inicialment, i el capítol 5 descriu els resultats més rellevants obtinguts al llarg de tot l'anàlisi. D'altra banda, el contingut i la discussió dels diferents resultats generats es precisa al capítol 6.

Seguidament, el capítol 7 resumeix les conclusions obtingudes a partir de la consecució del projecte presentat, i al capítol 8 es defineixen els diferents acrònims emprats en un glossari. Per últim, el capítol 9 inclou les referències bibliogràfiques utilitzades durant la realització la memòria.

## 3 Estat de l'art

**3.1 La microbiota intestinal de *C. elegans***

Actualment, l'examinació de les interaccions entre el nematode *C. elegans* i els microorganismes presents al seu entorn ha estat abordada per un reduït conjunt d'estudis, tot i les facilitats que presenta el metazou per

realitzar aquest tipus d'anàlisis experimentals <sup>[26]</sup>. No obstant, el recent interès en la caracterització del microbioma i les seves implicacions, en conjunt amb els avenços desenvolupats en les tecnologies de seqüenciació, ha propiciat l'emergència de diverses investigacions destinades al coneixement de la microbiota natural de *C. elegans* i altres espècies del gènere *Caenorhabditis* mitjançant la seqüenciació del gen 16S rDNA procariota <sup>[26,45]</sup>.

En general, la microbiota intestinal de *C. elegans* en l'entorn silvestre és molt diversa però estable geogràficament. De manera similar als mamífers i altres animals, la comunitat entèrica del nematode sembla estar constituïda per microorganismes dels fílums *Firmicutes*, *Bacteroidetes*, *Actinobacteria* i *Proteobacteria*, sent aquest últim el més predominant donada l'elevada presència de membres de la família *Enterobacteriaceae* <sup>[26,46]</sup> i també de les famílies *Pseudomonadaceae*, *Xanthomonadaceae* i *Sphingomonadaceae* <sup>[28,47,48]</sup>. Tot i això, els estudis apunten a l'existència de variacions considerables de la comunitat bacteriana intraespecífica induïda per l'ambient i la genètica de l'hoste. Per exemple, la microbiota intestinal d'individus de *C. elegans* sembla diferir de la present en el seu medi natural i també de la que presenta *C. remanei*, mentre que es correlaciona amb el perfil microbià característic de *C. briggsae* <sup>[26,46,49]</sup>. De fet, les comunitats microbianes semblen esdevenir més similars entre nematodes de la mateixa soca que entre soques o espècies diferents, i també solen ser semblants entre cucs de la mateixa espècie recollits de diferents substrats <sup>[28,49]</sup>. D'altra banda, la comparació entre nematodes extrets de l'ambient natural en contraposició amb altres enriquits al laboratori amb un microbioma experimental demostra algunes diferències en les seves comunitats entèriques, per la qual cosa l'entorn també pot influir en la microbiota intestinal dels cucs <sup>[48]</sup>. No obstant, altres observacions suggereixen que *C. elegans* mostra un microbioma característic que es defineix per les propietats que presenta l'espècie, ja que les comunitats d'individus extrets de diferents entorns presenten certes similituds significatives <sup>[26,49]</sup>.

Pel que fa a l'efecte de la composició de la microbiota entèrica sobre l'hoste, algunes investigacions indiquen que la composició i diversitat de la comunitat bacteriana ambiental es correlaciona amb les etapes de desenvolupament dels nematodes <sup>[26,45]</sup>. Concretament, els entorns rics en *Alphaproteobacteria* afavoreixen la proliferació dels cucs, mentre que nivells alts de *Gammaproteobacteria* i *Bacteroidetes* promouen l'aturada del seu creixement (fase d'enquistament) <sup>[50]</sup>. A més, els microorganismes del fílum *Proteobacteria*, especialment dels gèneres *Pseudomonas*, *Achromobacter*, *Acinetobacter* i *Comamonas*, s'associen a una taxa de reproducció i desenvolupament major respecte les condicions control, afavorint la *fitness* dels nematodes <sup>[26,49]</sup>. En general, la composició de la comunitat intestinal de *C. elegans* pot influenciar la seva immunitat, la germinació de cèl·lules germinals, la velocitat de creixement i la resistència a condicions d'estrès, entre d'altres <sup>[26,45,46]</sup>. Per tant, analitzar les interaccions hoste-microbioma a partir d'un model simple com el de *C. elegans* resulta útil per caracteritzar els efectes de diversos factors microbianes i les rutes de l'hoste en múltiples processos biològics <sup>[45]</sup>.

### 3.2 Mètodes d'estudi del comportament motor dels nematodes

Tal i com s'ha comentat anteriorment a l'apartat 2.1, *C. elegans* presenta certes característiques útils per tal d'avaluar els mecanismes intrínsecs de la locomoció animal. Aquest nematode respon a un ampli rang d'estímuls químics, mecànics i tèrmics, i s'ha utilitzat extensivament en diverses investigacions per tal d'aprofundir en els mecanismes moleculars que defineixen la recepció d'aquests paràmetres <sup>[51]</sup>. La resposta conductual del metazou a l'entorn es manifesta a partir d'alteracions en la seva locomoció, les quals han de ser analitzades mitjançant eines computacionals i matemàtiques adients <sup>[51]</sup>.

Actualment existeixen diverses aproximacions i mètodes dissenyats especialment per tal d'avaluar els patrons de moviment dels nematodes en un sistema de *tracking* <sup>[51,52,53]</sup>. Encara que molts dels algorismes i aplicacions

subjacents es basen en idees similars, fent-los conceptualment similars, existeixen diferències crítiques en la seva implementació, rendiment i espai d'aplicació <sup>[52]</sup>.

Diversos programes d'enregistrament artificial segueixen una estratègia similar de processament de dades, que implica inicialment l'extracció de les imatges individuals de cada fotograma a partir de l'arxiu de vídeo generat en el *tracking* <sup>[53,54]</sup>. A continuació, s'extrau la forma del cuc mitjançant l'assignació dels píxels corresponents al cuc i al fons, segons si la seva intensitat supera un cert llindar establert. Aquest procés genera imatges binàries en blanc i negre que s'analitzen per tal de discernir la columna vertebral dels nematodes, en un procés denominat esqueletització o *skeletonization* en anglès <sup>[54]</sup>. Així doncs, l'examinació de les imatges obtingudes dona lloc al càlcul de diversos paràmetres, com per exemple el centre de massa del cuc (anomenat com centroide) o els angles entre segments adjacents dels individus que mesuren la curvatura corporal <sup>[53,54]</sup>. En general, la velocitat dels nematodes es defineix com la taxa de canvi en la ubicació del seu centroide al llarg de l'esquelet en funció del temps, i es mesura a través de la seqüència de fotogrames individuals de la gravació <sup>[54]</sup>. No obstant, la disposició dels nematodes sota una estructura de seguiment en absència d'aliment provoca que aquests animals tendeixin a realitzar una exploració àmplia de l'àrea amb multitud de girs i inversions a gran velocitat, per la qual cosa l'examinació de la locomoció animal a partir dels sistemes mencionats encara suposa un repte avui dia <sup>[53,55]</sup>.

En resum, l'anàlisi del comportament de *C. elegans* i altres nematodes basat en el seguiment dels seus moviments sobre plaques d'agar resulta extremadament útil en l'estudi del comportament sensorial, la locomoció i l'efecte de diverses mutacions sobre el fenotip dels cucs <sup>[52,54,56]</sup>. Tot i això, l'efecte de la microbiota intestinal dels nematodes sobre els seus patrons motors encara no ha estat examinat mitjançant aquests mètodes, tot i que sí s'han dut a terme algunes aproximacions relacionades en altres organismes model com *Drosophila* o *Mus musculus* <sup>[57,58,59]</sup>. Per tant, la caracterització diferencial de la microbiota intestinal i de la locomoció de diferents soques de *C. elegans* de manera independent suposa una base per a futures investigacions dedicades a l'estudi de la interconnexió existent entre aquests dos àmbits.

## 4 Metodologia

En aquest projecte s'exposen dos blocs d'estudi centrats en l'organisme model *C. elegans*, en cadascun dels quals s'implementa una metodologia d'anàlisi pròpia i independent de l'altra. Resulta convenient, doncs, separar ambdós esquemes de treball per oferir una visió organitzada de les dues aproximacions procedimentals seguides. El *scripts* emprats poden consultar-se al repositori GitLab pertinent (veure Annex III).

### 4.1 Caracterització de la microbiota intestinal del nematode *C. elegans*

En aquesta primera part de l'estudi s'ha elaborat un codi executable i reproduïble útil per a la descripció de la comunitat microbiana intestinal de diferents nematodes. Amb el *pipeline* bioinformàtic definit, que combina llenguatge GNU-Bash i R, s'ha dut a terme un meta-anàlisi amb vuit *datasets* de seqüències que corresponien a metagenomes d'amplicó del gen *16S* de la microbiota entèrica dels cucs. Aquests es troben descrits a la literatura i es poden descarregar des dels repositoris públics NCBI – SRA (*Sequence Read Archive*), MG-RAST (*Metagenomic Rapid Annotations using Subsystems Technology*) i EBI – ENA (*European Bioinformatics Institute – European Nucleotide Archive*), segons el cas.

El pre-processat inicial de la *raw data* amb llenguatge GNU-Bash s'ha executat de forma remota amb el clúster de computadores del CEAB-CSIC, el qual funciona amb Bio-Linux8 (Ubuntu 14.04LTS) i disposa de 13 nodes i 284 processadors. Els usuaris de Windows poden accedir a la terminal del clúster mitjançant PuTTY (v.0.76) <sup>[60]</sup>, una interfície de treball que a través de xarxa permet iniciar sessió en una computadora multiusuari des d'un

altre ordinador. D'altra banda, l'anàlisi descriptiu dels resultats obtinguts en el pre-processat s'ha dut a terme de forma local amb R (v.4.1.3) <sup>[42]</sup> i RStudio (v.2022.02.0+443) <sup>[43]</sup>. Tots els paquets emprats en l'estudi s'han instal·lat i actualitzat en la seva darrera versió per evitar problemes d'incompatibilitat.

#### 4.1.1 Mètodes i procediments del meta-anàlisi

##### A. Descàrrega i descripció inicial dels *datasets* de seqüències

El primer pas va consistir en recopilar, mitjançant el motor de cerca PubMed de NCBI, la informació disponible envers la descripció de la microbiota intestinal de *C. elegans* i altres tàxons filogenèticament propers (com soques de la mateixa espècie o, fins i tot, altres espècies del gènere *Caenorhabditis*). Algunes de les paraules clau emprades en aquesta cerca foren “Caenorhabditis”, “elegans”, “assembly”, “microbiota”, “natural”, “native”, “microbiome”, “gut”, “intestinal”, “gut-brain axis” i “animal model”. Com a resultat, s'identificaren sis articles de referència publicats en aquest àmbit, cadascun dels quals portava associat un o més *datasets* de seqüències accessibles des de diferents repositoris públics.

Una revisió exhaustiva d'aquests sis articles fou necessària, doncs comprendre les aproximacions d'estudi i les estratègies experimentals proposades pels autors originals marcaria el rumb de les comparatives a realitzar entre els diversos *datasets* (re-nomenats a conveniència per a una manipulació més còmode de les dades). A continuació es descriu breument el propòsit de cadascun dels sis projectes seleccionats:

1. Article *Assembly of the Caenorhabditis elegans gut microbiota from diverse soil microbial environments*, de Maureen Berg <sup>[28]</sup>: en aquest estudi es pretenia caracteritzar el microbioma entèric natiu de cucs *C. elegans* soca N2, mantinguts en un ambient de microcosmos amb restes de sòl i podridures que emulaven l'hàbitat natural d'aquest nematode. El metagenoma d'amplicó associat (re-anomenat com **Berg\_assembly**) es troba emmagatzemat al repositori MG-RAST sota l'identificador *mgp13213*.
2. Article *Host-Specific Functional Significance of Caenorhabditis Gut Commensals*, de Maureen Berg <sup>[47]</sup>: en aquest cas es va examinar la contribució de la genètica de l'hoste en la composició de la microbiota intestinal, identificant diferències entre les comunitats natives de cucs de diferents genotips. Aquí s'incloueren 4 espècies de *Caenorhabditis* (*C. elegans*, *C. briggsae*, *C. remanei* i *C. tropicalis*) i 3 soques de *C. elegans* (N2, Hawaiian (o CB4856) i CB4857). També s'afegí un *outgroup*, *Pristionchus pacificus*. Igual que abans, tots aquests nematodes foren mantinguts en condicions de microcosmos, aprofitant fraccions de sòl i podridures de l'entorn natural. El *dataset* associat (re-anomenat com **Berg\_host**) també es troba disponible al repositori MG-RAST, en aquest cas sota l'identificador *mgp18910*.
3. Article *The native microbiome of the nematode Caenorhabditis elegans: gateway to a new host-microbiome model*, de Philipp Dirksen <sup>[49]</sup>: aquest projecte es desdoblava en dues parts. La primera d'elles es centrava en descriure la microbiota intestinal natural de nematodes salvatges: uns aïllats i processats directament sense exposició a *Escherichia coli* OP50 <sup>1</sup> (concretament *C. elegans*, *C. briggsae* i *C. remanei*), i uns altres aïllats i processats després de dues setmanes d'exposició en placa a aquest bacteri (només *C. elegans* i *C. briggsae*). Les

<sup>1</sup> Soca bacteriana que habitualment serveix d'aliment pel manteniment dels nematodes al laboratori (veure apartat B de la secció 4.2.1 del present TFM).

lectures crues vinculades a aquest *dataset* (identificat com **Dirksen\_native**) poden descarregar-se des de l'entrada PRJEB12994 de la base de dades EBI-ENA.

A la segona part es re-infectaven soques axèniques N2, MY379 i MY316 de *C. elegans* amb una comunitat microbiana experimental, constituïda per 14 tàxons bacterians representatius de la microbiota intestinal natural de la mateixa espècie. L'objectiu era comprovar l'habilitat colonitzadora dels diferents membres i la influència que tenien sobre el creixement del cuc. El *dataset* corresponent (re-nombrat com **Dirksen\_experimental**) es troba emmagatzemat a la base de dades EBI-ENA sota l'identificador PRJEB13031.

4. Article *CeMbio - The Caenorhabditis elegans Microbiome Resource*, de Philipp Dirksen <sup>[61]</sup>: els autors varen definir una comunitat microbiana experimental model formada per 12 tàxons bacterians, adequada per a la colonització intestinal de cucs axènics de l'espècie *C. elegans*. Aquests tàxons varen ser seleccionats en base a tres criteris: 1) havien de tenir significança ecològica, és a dir, trobar-se a la microbiota nativa del nematode, o com a mínim al seu hàbitat natural; 2) havien de recompondre l'àmplia diversitat de famílies taxonòmiques d'aquesta comunitat natural; i 3) havien de ser fàcils de cultivar i mantenir en medis estàndard. Es portaren a terme dues proves de colonització per separat, cadascuna a un laboratori diferent i amb una metodologia de treball similar, però variant reactius i medis de cultiu, el moment de mostreig dels 12 membres bacterians i les soques de *C. elegans* manipulades (N2 en ambdós experiments, CB4856 només en un d'ells). Els metagenomes d'amplicó associats a cada experiment, amb les seqüències crues corresponents, es troben disponibles a la base de dades EBI-ENA sota els identificadors PRJEB37101 i PRJEB37035 (re-nombrats com **DirksenCeMbio\_exp2** i **DirksenCeMbio\_exp3**, respectivament).
5. Article *Community assembly of the native C. elegans microbiome is influenced by time, substrate and individual bacterial taxa*, de Julia Johnke <sup>[48]</sup>: aquest grup d'investigació va ressaltar la importància del substrat i l'estadi de desenvolupament del metazou en la composició nativa de la microbiota intestinal de cucs salvatges de l'espècie *C. elegans*. Alguns individus foren exposats a *E. coli* OP50 després de l'aïllament, mentre que la resta foren aïllats i processats directament sense exposició al bacteri. El *dataset* vinculat (identificat amb el cognom de l'autora, **Johnke**) es troba emmagatzemat a l'entrada PRJEB33740 del repositori EBI-ENA.
6. Article *Caenorhabditis elegans responses to bacteria from its natural habitats*, de Buck Samuel <sup>[50]</sup>: en aquest últim estudi es caracteritzà la microbiota pròpia de l'hàbitat natural de *C. elegans*. Per tant, no es processaren nematodes sinó fraccions de substrat on s'havia confirmat la presència d'aquests cucs, com podridures vegetals i peces de fruita en procés de descomposició. Aquest *dataset* (identificat amb el cognom de l'autor, **Samuel**) es pot descarregar des de l'entrada SRS1849345 de la base de dades NCBI-SRA.

Tal i com es pot comprovar, cada projecte es centrava en caracteritzar el perfil microbià intestinal de diferents espècies i/o soques de nematodes que havien proliferat en determinades condicions de creixement. Així doncs, amb els vuit *datasets* mencionats es va plantejar la realització d'un meta-anàlisi comparatiu de la comunitat microbiana entèrica dels membres del gènere *Caenorhabditis* entre les diferents condicions ambientals establertes. Aquestes condicions de creixement es valoraren dins d'un gradient de naturalització, donat que els hostes procedien de distints entorns de desenvolupament: alguns cucs eren silvestres, altres foren aïllats i enriquits al laboratori o bé disposats en microcosmos, i

uns últims es mantingueren en condicions semi-artificials i completament experimentals. En qualsevol cas, l'espècie *C. elegans* serviria sempre d'eix conductor en totes les comparatives a efectuar.

Els *datasets* emmagatzemats a MG-RAST i NCBI-SRA es varen descarregar amb el paquet de funcions `grabseqs` v.0.7.0, el qual està implementat en Python, és compatible amb la majoria de sistemes operatius de Linux (com Ubuntu) i pot instal·lar-se a través de Conda per tal d'activar automàticament les diferents dependències del programari [62]. El paquet en qüestió permet a l'usuari baixar tots els *reads* de totes les mostres associades a un identificador de projecte per mitjà de les funcions `grabseqs sra` o `grabseqs mgrast`, que s'executen des de la terminal del clúster amb un *script* de GNU-Bash [62]. Per emprar aquestes comandes cal especificar el directori de destí (una carpeta amb el nom curt assignat al metagenoma d'amplicó corresponent), el nº de *threads* (la potència de descàrrega) i l'ID del repositori públic des d'on es volen abaixar les dades [62]. Dit això, les lectures de seqüenciació dels *datasets Berg\_assembly* i *Samuel* varen guardar-se amb extensió FASTQ.gz. No es varen diferenciar els arxius comprimits en R1 i R2, donat que el primer dels metagenomes mencionats estava directament dipositat en format *merged* i el segon estava constituït per *single-end reads*. Per un altre costat, les lectures de *Berg\_host* només foren accessibles en format FASTA.

D'altra banda, els *datasets* apujats a EBI-ENA foren descarregats localment des de la pròpia entrada a la base de dades, fent *click* a l'opció *Submitted files FTP*. Els *reads* de cada metagenoma es varen guardar amb extensió FASTQ.gz, distingint ara sí les lectures aparellades en dos arxius comprimits R1 i R2. Totes aquestes seqüències foren apujades després al clúster mitjançant l'eina FileZilla [63].

Paral·lelament, i a partir dels arxius de metadata originals de cada estudi, també fou necessari crear un fitxer de metadata global per integrar la informació sobre els diferents experiments de seqüenciació. Per a cada *dataset* es varen descriure les següents categories: *Sample\_ID* (els identificadors de les mostres), *Author* (l'autor de correspondència), *BioProject* (el codi identificador del projecte a la base de dades), *Nematoda\_Genera/Species/Strain* (el gènere, l'espècie i la soca del nematode processat), *Material\_Feature* (el substrat de partida), *Sequencing\_Method* i *Instrument* (el mètode i la plataforma de seqüenciació, principalment Illumina MiSeq/HiSeq), *Primers* (els encebadors emprats en el procés de seqüenciació), *16S\_Target\_subfragment* (el fragment diana del gen *16S*, que en general engloba les regions V3-V4) i *Location\_Sample* (el lloc de mostreig).

## B. Pre-processat inicial dels *datasets*: *merging*

Els *reads* descarregats no havien de passar per una fase prèvia de *demultiplexing* dels *barcodes*, perquè les seqüències estaven ja assignades a les diferents mostres problema. Com a conseqüència, el tractament inicial de les lectures s'inicià amb una fase de *merging*. Es tracta d'un procediment bioinformàtic que permet concatenar els arxius R1 i R2, prèviament descomprimits, associats a un mateix grup de seqüències d'extremes aparellats [64]. Com a resultat, s'obté un fragment coincident o consens entre les lectures complementàries, derivat de l'encavalcament entre els *reads* obtinguts en ambdues direccions de seqüenciació [64].

Cal remarcar que tant el *merging* com l'anàlisi de qualitat i el filtratge de seqüències (que es detallen en apartats posteriors) es varen aplicar sobre cada *dataset* per separat i a través de *scripts* implementats en llenguatge GNU-Bash, ja que es pretenia inferir la qualitat particular dels *reads* de cada metagenoma per, conseqüentment, establir els criteris de cribratge pertinents.

Per dur a terme el *merging* es va emprar l'algoritme de USEARCH [39], una eina de processament i anàlisi que, entre d'altres coses, es caracteritza per assignar seqüències a clústers i cercar coincidències d'alta

identitat (*hits*) amb una rapidesa de diversos ordres de magnitud superior a BLAST <sup>[65]</sup>. Específicament, les lectures aparellades foren assemblades en un únic arxiu d'extensió FASTQ mitjançant la comanda `fastq_mergepairs` del programa USEARCH v.11.0.667 <sup>[66]</sup>. En la funció cal determinar el màxim nº de *mismatches* a admetre en l'alineament (argument `fastq_maxdiffs`) i el llindar mínim d'encavalcament requerit per sota del qual descartar concatenats (argument `fastq_minovlen`); de forma suplementària, es poden fixar els recursos de CPU necessaris per al còmput (argument `threads`) i sol·licitar la sortida d'un fitxer .txt amb el *report* final de l'execució (argument `report`) <sup>[66]</sup>.

En el cas dels *datasets Berg\_assembly* i *Samuel*, el *merging* es realitzà amb la comanda bàsica de GNU-Bash `cat`, la qual va permetre concatenar les lectures prèviament descomprimides de les diferents mostres disponibles en un únic arxiu d'extensió FASTQ. De forma similar, els *reads* del *dataset Berg\_host* es varen assemblar amb la mateixa funció `cat` en un sol fitxer de format FASTA.

Per últim, es va comprovar que els identificadors de les mostres de cada *dataset* estiguessin formatejats correctament amb la comanda `fastx_get_sample_names`, també de USEARCH v.11.0.667 <sup>[67]</sup>. Aquesta funció crea un *report* en format .txt amb els IDs de les mostres avaluades, que idealment han de coincidir sempre amb els presentats a la metadata.

### C. Pre-processat inicial dels *datasets*: anàlisi de qualitat

Amb les lectures concatenades procedents del pas anterior va ser possible estudiar la qualitat global de cada *dataset* individualment. Aquesta tasca s'implementà a través d'un nou *script* de llenguatge GNU-Bash, emprant en aquest cas l'algoritme de VSEARCH <sup>[40]</sup>. Es tracta d'una alternativa a USEARCH de codi obert i d'alta qualitat que incorpora la major part de les seves funcions i també de noves, però no té limitació de memòria RAM i admet fitxers més grans de 4Gb (cosa que USEARCH no pot fer) <sup>[40]</sup>. Resulta especialment útil quan es volen manipular arxius de gran volum, els quals no podrien ésser processats de forma tradicional per USEARCH <sup>[40]</sup>. El codi font d'aquest últim programa no es troba disponible públicament, i de forma gratuïta només s'hi pot accedir a la versió acadèmica de 32-bit; VSEARCH es va desenvolupar per suplir totes aquestes mancances, i és tant o més ràpid i precís que USEARCH en la majoria de tasques bioinformàtiques (incloses el *clustering* i l'alineament de seqüències i quimeres) <sup>[40]</sup>.

Concretament, l'anàlisi de qualitat es portà a terme amb la comanda `fastq_stats` de VSEARCH v.2.13.4 <sup>[68]</sup>, amb la qual s'obtingueren les estadístiques necessàries per identificar la longitud òptima de truncament dels *reads* i l'*expected error* amb què filtrar-los. Aquesta funció pren com *input* les seqüències *merged* del metagenoma en qüestió i torna com *output* un arxiu d'extensió .log (que pot convertir-se a .txt) amb les estadístiques mencionades, associades principalment a la distribució de la longitud dels *reads*, els *expected errors* i els índexs de qualitat <sup>[68]</sup>.

El fitxer d'estadístiques generat es va descarregar de forma local per tal de representar amb R els valors de qualitat computats, amb l'objectiu de determinar els dos criteris de filtratge a implementar en el pas següent. A continuació, es mostren a mode d'exemple els gràfics obtinguts per a un dels *datasets* en particular, *Dirksen\_native*, i la seva corresponent interpretació. L'anàlisi de qualitat dels altres *datasets* pot consultar-se a l'Annex I.

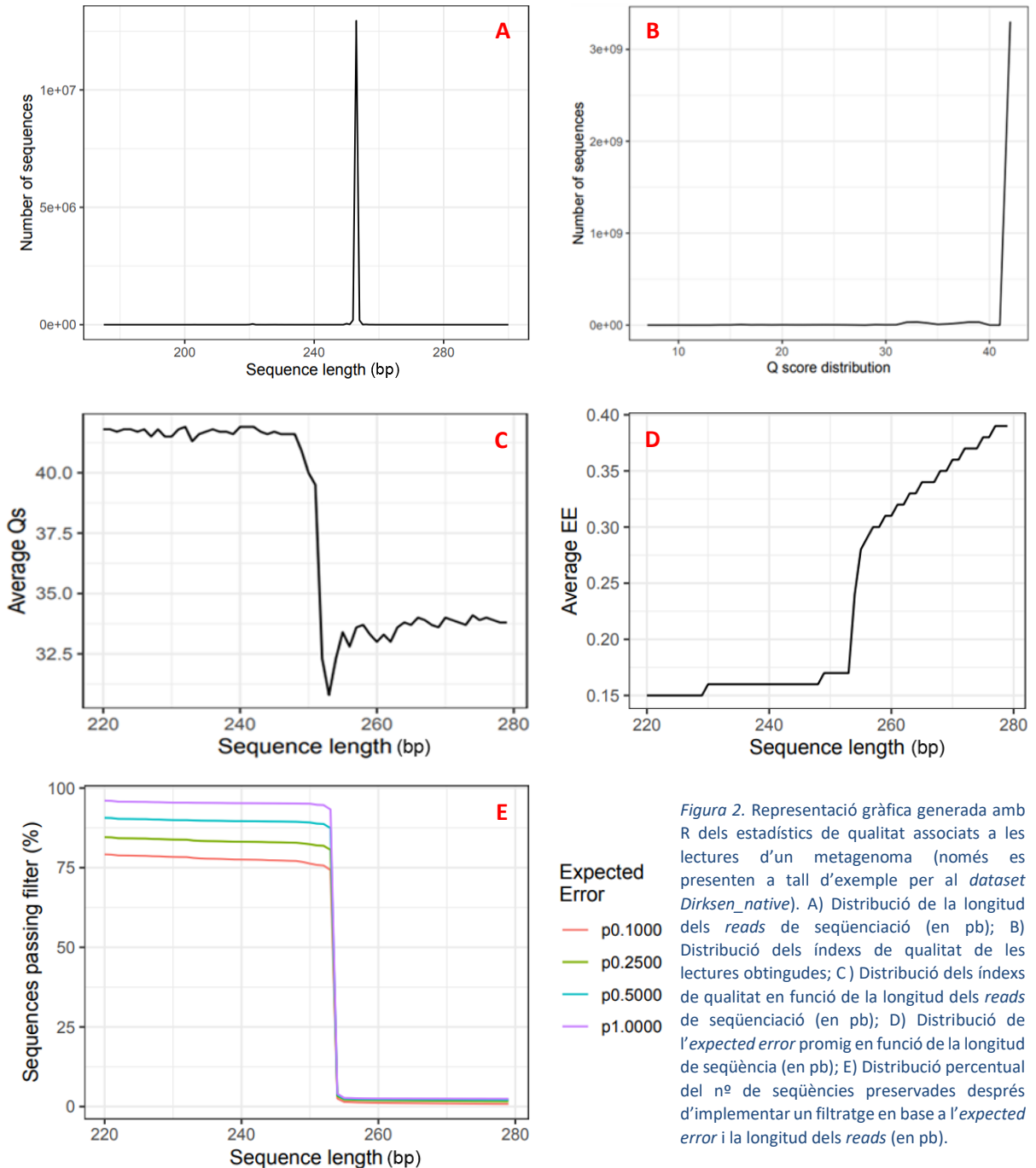


Figura 2. Representació gràfica generada amb R dels estadístics de qualitat associats a les lectures d'un metagenoma (només es presenten a tall d'exemple per al *dataset Dirksen\_native*). A) Distribució de la longitud dels *reads* de seqüenciació (en pb); B) Distribució dels índexs de qualitat de les lectures obtingudes; C) Distribució dels índexs de qualitat en funció de la longitud dels *reads* de seqüenciació (en pb); D) Distribució de l'*expected error* promig en funció de la longitud de seqüència (en pb); E) Distribució percentual del nº de seqüències preservades després d'implementar un filtratge en base a l'*expected error* i la longitud dels *reads* (en pb).

Tal i com es pot comprovar, la major part dels *reads* del *dataset Dirksen\_native* tenen una longitud aproximada de 250 pb (Fig. 2A) i un índex de qualitat superior a 40 (Fig. 2B). Aquests índexs, també coneguts com *Q-scores*, descriuen una mètrica sobre la precisió del procés de seqüenciació, i indiquen la probabilitat estimada amb què una base nucleotídica s'ha anotat incorrectament<sup>[69]</sup>. Es calculen com  $-10\log_{10}(P)$  (on  $P$  és la probabilitat d'error en l'anotació d'una base) i acostumen a oscil·lar entre 0 i 50<sup>[69]</sup>. Això significa que, en el cas de *Dirksen\_native*, amb una precisió de seqüenciació del 99.99% la



probabilitat de trobar una base errònia és d'1 entre 10,000. Resulta remarcable la disminució del *Q-score* promig per sota de 40 quan les longituds dels *reads* són superiors a 250 pb (Fig. 2C).

En canvi, l'*expected error* (EE) és una mesura promig dels errors que es detecten en una àmplia col·lecció de seqüències en les quals la taxa d'error per posició ve definida pel *Q-score*, assumint que els errors en posicions diferents es produeixen de forma independent <sup>[70]</sup>. Tenint això en compte, es veu clarament que la mitjana d'EE de les lectures sembla mantenir-se al voltant de 0.175 fins els 250 pb, i a longituds superiors es produeix un augment dràstic d'aquest valor (Fig. 2D). A més a més, considerant qualsevol dels llindars plausibles per a l'EE és possible preservar més del 75% de les seqüències del *dataset* amb 250 pb de longitud (Fig. 2E).

Val a dir que no es va poder executar l'anàlisi de qualitat sobre el metagenoma *Berg\_host*, donat que la comanda `fastq_stats` de VSEARCH treballa només amb *inputs* en format FASTQ i aquest *dataset* estava constituït per lectures amb extensió FASTA.

#### D. Pre-processat inicial dels *datasets*: filtratge de seqüències

El següent pas del meta-anàlisi va consistir a cribrar les lectures dels vuit metagenomes per separat amb l'objectiu de conservar les seqüències amb un major grau de qualitat. En base als gràfics generats a l'apartat anterior, es varen estipular els següents criteris de filtratge específics per a cadascun dels *datasets* d'estudi:

- Metagenoma d'amplicó *Berg\_assembly*: truncament dels *reads* a **150 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *Berg\_host*: truncament dels *reads* a **150 pb**.
- Metagenoma d'amplicó *Dirksen\_native*: truncament dels *reads* a **250 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *Dirksen\_experimental*: truncament dels *reads* a **250 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *DirksenCeMbio\_exp2*: truncament dels *reads* a **250 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *DirksenCeMbio\_exp3*: truncament dels *reads* a **400 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *Johnke*: truncament dels *reads* a **450 pb** i llindar d'EE de **0.25**.
- Metagenoma d'amplicó *Samuel*: truncament dels *reads* a **245 pb** i llindar d'EE de **0.25**.

El *dataset* *Berg\_host* només es va poder depurar en base a la longitud dels *reads*, ja que no es disposava dels estadístics de qualitat pertinents per valorar el llindar d'*expected error* necessari pel cribratge. La longitud de truncament especificada es va raonar a partir del mètode de seqüenciació emprat en l'estudi original (150-bp Illumina HiSeq *paired-end reads*) <sup>[47]</sup>.

El filtratge com a tal es va executar amb la comanda `fastq_filter` de VSEARCH v.2.13.4 <sup>[68]</sup> a través d'un *script* de llenguatge GNU-Bash. Aquesta funció pren com *input* les seqüències *merged* del metagenoma en qüestió i retorna com *output* les lectures filtrades en format FASTA <sup>[68]</sup>. El cribratge s'aplica per mitjà dels arguments `fastq_maxee` i `fastq_truncLen`, que defineixen el llindar màxim permès per a l'EE de les seqüències i la longitud a la qual escindir-les, respectivament <sup>[68]</sup>. Si els *reads* tenen una longitud inferior al valor de truncament, aleshores es descarten per complet <sup>[68]</sup>. De forma addicional, es pot emprar l'argument `fastq_qmax` per especificar el *Q-score* màxim admès pels diferents fitxers FASTQ a processar (per exemple  $Q_s = 42$ ) <sup>[68]</sup>.

### E. Pre-processat conjunt dels *datasets*: identificació de seqüències úniques

Arribats a aquest punt, els vuit *datasets* filtrats es varen concatenar amb la comanda `cat` en un únic arxiu FASTA. Això facilitaria el tractament simultani de totes les lectures en els posteriors processos d'assignació a OTUs i classificació taxonòmica (apartats F i G).

Ara bé, abans de prosseguir amb aquestes activitats s'havia de realitzar una neteja del meta-*dataset*. La finalitat d'aquesta etapa procedimental era, en essència, l'eliminació de seqüències duplicades. És a dir, s'havien d'identificar els *reads* únics a partir de la preservació d'una sola còpia de cada lectura.

En primer lloc, fou necessari ordenar les seqüències per longitud decreixent per tal d'optimitzar la detecció d'aquelles que esdevenien idèntiques. Aquesta tasca s'implementà a través d'un nou *script* de llenguatge GNU-Bash, emprant en aquest cas la comanda `sortbylength` de VSEARCH v.2.13.4 que requereix com *input* el fitxer de *datasets* filtrats i concatenats <sup>[68]</sup>.

En segon lloc, es varen col·lapsar les seqüències estrictament idèntiques (amb la mateixa longitud i successió de nucleòtids) per preservar només una de les còpies <sup>[68]</sup>. Això es va assolir amb la funció `derep_fulllength` de VSEARCH v.2.13.4, implementada en un nou *script* de llenguatge GNU-Bash <sup>[68]</sup>. La funció pren per *input* l'arxiu de *datasets* filtrats, concatenats i ordenats prèviament, i en retorna un altre només amb les lectures úniques <sup>[68]</sup>. S'incloueren els arguments `sizeout` (per anotar valors d'abundància als encapçalaments FASTA), `fasta_width` (que establert a zero permet mantenir tota la seqüència FASTA en una sola línia de fitxer) i `minuniquesize` (per descartar les seqüències col·lapsades amb una abundància menor al valor estipulat, en aquest cas 2) <sup>[68]</sup>. Els *reads* conservats foren ordenats una segona vegada amb `sortbylength`.

Seguidament, es va creure convenient eliminar totes aquelles seqüències que es trobaven de forma íntegra en altres de major longitud (sub-seqüències idèntiques) <sup>[71]</sup>. En aquest cas, el col·lapse s'executà amb l'opció `derep_subseq` de la comanda `cluster` de USEARCH v.5.2.32, aplicada en un nou *script* de llenguatge GNU-Bash <sup>[71]</sup>. La funció normalment s'inicialitza amb els *reads* conservats i ordenats d'abans, i retorna un fitxer amb les lectures úniques definitives. S'especificaren els arguments `sizein` i `sizeout` (per calcular nous valors d'abundància), `slots` (per indicar els recursos necessaris de memòria computacional), `minlen` (per fixar la longitud de cada sub-seqüència o *query* a cercar, en aquest cas de 64 nt) i `w` (per establir la longitud mínima de coincidència entre la *query* i la resta de lectures a 32 nt) <sup>[72]</sup>. Per defecte, `derep_subseq` marca l'argument `dbstep` a 1, assegurant així que la finestra de desplaçament entre els intervals de cerca de 32 nt sigui d'una unitat i, per tant, s'avaluï la totalitat de les seqüències nucleotídiques <sup>[72]</sup>.

Per últim, les lectures úniques obtingudes foren ordenades per abundància decreixent amb la funció `sortbysize` de USEARCH v.11.0.667, executada des d'un nou fitxer de codi redactat en llenguatge GNU-Bash <sup>[73]</sup>. Com era d'esperar, la funció utilitza com *input* les seqüències de-replicades del pas anterior, tot especificant l'argument `minsize` per descartar les lectures col·lapsades amb una abundància menor al valor estipulat, en aquest cas de 2 <sup>[73]</sup>.

### F. Pre-processat conjunt dels *datasets*: generació d'OTUs al 97% d'identitat

El concepte d'OTU fou encunyat als anys 60 per Peter Sneath i Robert Sokal, i responia a la necessitat de desenvolupar una estratègia quantitativa per classificar els organismes en funció dels seus caràcters observables <sup>[74,75]</sup>. Els autors pretenien crear una categorització jeràrquica que representés de la forma més fidedigne possible les relacions filogenètiques entre éssers vius <sup>[74,75]</sup>, tenint en compte que llavors

les seqüències de DNA eren totalment inaccessibles i no s'havien dissenyat encara algorismes de reconstrucció filogenètica (com *neighbor-joining* o màxima versemblança) [75].

El terme en qüestió seria revitalitzat per la comunitat científica dècades més tard, als anys 90, època en què no resultava factible caracteritzar noves lectures del gen *16S* (obtingudes per mètode Sanger) a través de cerques a les bases de dades públiques, que per aleshores eren escasses i pobres [75]. Es va idear una alternativa més pragmàtica centrada en la generació de clústers de seqüències al 97% d'identitat, els quals podien ser tractats com OTUs de Sneath i Sokal per assignar de forma temptativa aquestes seqüències a espècies i/o altres rangs taxonòmics superiors [75]. Tot i que l'aproximació era raonable, només tenia validesa si els errors de lectura eren menyspreables [39,75]: les tècniques de seqüenciació de 1<sup>a</sup> generació, com per exemple Sanger, presentaven una taxa d'error ínfima i s'ajustaven adequadament a aquest procediment [76], però les posteriors tecnologies de *Next-Generation Sequencing* (NGS) varen fer notòria la necessitat d'implementar mètodes de pre-processat dels *reads* per disminuir aquesta taxa i optimitzar l'adjudicació correcta dels OTUs (descartant en el procés quimeres i lectures espúries) [75,77]. Avui dia, els OTUs es defineixen com grups de seqüències que, de manera aproximada, pretenen especificar un clade taxonòmic o un grup monofilètic concret, generalment una espècie.

L'algoritme de UPARSE es troba integrat al programa USEARCH [39] i permet computar el conjunt d'OTUs més representatius emprant un llindar mínim d'identitat del 97%, reportant un percentatge de bases errònies inferior al que s'obtidria amb altres mètodes de *clustering* (com QIIME o mothur) [41,78]. El *software* retorna un **subset de totes les seqüències biològicament correctes** presents en els *reads* i, per tant, un menor nº d'OTUs que esdevé consistent amb el valor esperat d'espècies de la comunitat d'estudi [41,75]. Totes les parelles d'OTUs contemplades pel programari han de diferir en un 3% o més d'identitat, i aquelles que resulten quimèriques s'acaben descartant [79]. Si una lectura encaixa en dos o més OTUs, l'assignació es duu a terme per màxima similitud d'alineament; si l'empat persisteix, la designació s'aplica sobre el primer OTU contrincant que ha trobat en la col·lecció [41,75].

Dit això, la generació dels OTUs al 97% d'identitat es va realitzar amb la comanda `cluster_otus` de USEARCH v.11.0.667 implementada en llenguatge GNU-Bash, prenent per *input* les seqüències úniques ordenades per abundància procedents de l'apartat anterior [80]. En total s'obtingueren 55,795 OTUs diferents.

A continuació, es va generar la taula d'OTUs necessària per a l'anàlisi gràfic dels resultats en R. Una taula d'OTUs és una matriu que informa sobre el recompte de *reads* (o la freqüència d'aquests en una escala de 0 a 1) per mostra (columnes) i per OTU (files) [81]. Ara bé, l'abundància de les lectures no necessàriament es correlaciona amb l'abundància de les espècies; la interpretació és més complexa i cal considerar els biaixos introduïts durant la obtenció dels amplicons i la seva seqüenciació [82]. La taula en qüestió es va computar amb la funció `usearch_global` de VSEARCH v.2.13.4 mitjançant una comanda de llenguatge GNU-Bash [68]. Aquest executable compara a través d'un alineament global les seqüències d'OTUs, indicades amb l'argument `db`, amb les lectures FASTA dels *datasets* filtrats i concatenats sense col·lapsar, ja que d'aquesta manera es facilita l'associació de cada *read* amb la corresponent mostra i OTU [68]. La taula resultant, que queda designada amb l'argument `otutabout`, és una matriu de text pla que separa per tabuladors les abundàncies dels OTUs a les diferents mostres [68]. Altres arguments emprats foren `strand` (per comparar ambdues cadenes nucleotídiques en l'alineament) i `id` (per fixar el límit d'identitat al 97%) [68].

### G. Pre-processat conjunt dels *datasets*: classificació taxonòmica dels OTUs

L'últim pas del pre-processat tenia per objectiu vincular els OTUs de l'apartat anterior a una categoria taxonòmica específica. Per fer aquesta classificació es va emprar la base de dades d'RNA ribosomal SILVA, en particular la col·lecció *SSU Ref NR 99 138.1* [83]. Aquest *dataset* de referència està constituït per 510,508 seqüències representatives i no-redundants del gen *16S* rDNA, i fou creat amb un criteri del 99% d'identitat per tal de suprimir de la col·lecció original (*SSU Ref 138.1*) les cadenes nucleotídiques amb un elevat grau de similitud [83,84].

Cal matisar, però, que el repositori SILVA conté molts altres *datasets* de seqüències ribosomals que són revisats i actualitzats periòdicament, tant de la subunitat gran com petita [83]. De fet, hi ha dades emmagatzemades per als tres dominis de la vida (*Bacteria*, *Archaea* i *Eukarya*), i l'usuari pot descarregar la col·lecció pertinent segons l'estudi filogenètic que vulgui implementar [83].

Dit això, per accedir a la col·lecció de seqüències mencionada i assignar els OTUs als diferents tàxons va ser necessari utilitzar l'eina bioinformàtica SINA (acrònim de *SILVA Incremental Aligner*) [85]. Es tracta d'un executable dissenyat especialment per a la realització d'alineaments múltiples de seqüències de gens ribosomals, com els provinents del repositori SILVA [85]. També esdevé útil per fer una cerca per homologia basada en l'alineament computat i generar una classificació de les seqüències problema segons els resultats obtinguts [85]. Aquest programa pot inicialitzar-se a través d'un *script* de llenguatge GNU-Bash, però no hi ha un comandament específic per fer-ho. Cal precisar la ruta completa fins on es trobi el *software* i detallar els següents arguments:

- `i`: el fitxer d'entrada amb les seqüències a alinear i classificar, que en aquest cas foren els OTUs avaluats en l'apartat F [85].
- `o`: el nom que pren el fitxer de sortida amb les seqüències d'OTUs catalogades taxonòmicament [85].
- `intype` i `outtype`: format del fitxer d'entrada o sortida respectivament, que en ambdós casos fou FASTA [85].
- `ptdb`, `search` i `search-db`: directori on es troba la col·lecció de seqüències de referència (*Ref NR 99 138.1*) a emprar en l'alineament i la classificació de les lectures problema [85].
- `lca-fields`: configuració dels camps jeràrquics de classificació taxonòmica de SILVA fins l'anomenat *Last Common Ancestor* (LCA), la qual cosa permet anotar la filogènia de les seqüències *input* a partir de l'alineament d'aquestes amb el *dataset* de referència [85].
- `meta-fmt`: format de la metadata taxonòmica a exportar en cas que el fitxer de sortida amb les seqüències catalogades tingui una extensió FASTA [85]. En aquest meta-anàlisi es va optar per generar l'arxiu en qüestió en format .CSV.
- `search-min-sim`: valor mínim d'identitat a contemplar entre la cadena *input* i la seqüència de referència, fixat en aquest cas a 0.8 (un 80%) [85].
- `search-max-result`: màxim nº de *matches* a considerar per a cada seqüència *input*, delimitat en aquest cas a 10 [85].
- `lca-quorum`: llindar per al consens de la classificació taxonòmica, establert en aquest cas també a 0.8 [85]. Això significa que, per assignar un OTU a un rang taxonòmic determinat, calia que com a mínim 8 de les 10 seqüències de referència avaluades estiguessin classificades de la mateixa manera [85].

## H. Anàlisi descriptiva i estadística de les dades pre-processades

El meta-anàlisi va concloure amb un estudi gràfic de la composició de les mostres procedents del pas anterior, les lectures de les quals havien estat finalment catalogades a nivell taxonòmic. Per acomplir aquesta tasca fou necessari descarregar localment l'arxiu d'OTUs al 97% d'identitat, l'*OTU table* amb les abundàncies associades (veure apartat F) i l'assignació taxonòmica de cadascun d'ells (veure apartat G). La informació d'aquests tres documents es va adequar amb un editor de text per facilitar el posterior tractament de les dades, centrat en explorar de forma comparativa la microbiota intestinal dels membres del gènere *Caenorhabditis* entre les diferents condicions ambientals de creixement. Aquesta aproximació es va efectuar mitjançant la implementació d'un *script* de llenguatge R redactat en format RMarkdown.

D'entrada, es varen portar a terme diverses tasques preliminars a fi de reduir la complexitat del *meta-dataset*, entre elles l'associació de la taula d'OTUs amb les anotacions taxonòmiques corresponents i l'eliminació de seqüències no procariotes (és a dir, lectures identificades com cloroplasts, mitocondris o eucariotes). Addicionalment, es varen suprimir de la metadata i l'*OTU table* totes aquelles mostres amb menys de 3,000 lectures (considerades poc resolutives) i també les marcades com a control de l'estudi.

A continuació, es varen elaborar diferents gràfics descriptors per tal d'il·lustrar i contrastar la variada composició microbiana de les mostres de diferents procedències (és a dir, entre espècies de *Caenorhabditis* i també entre condicions de creixement). D'una banda, es va executar un escalat multidimensional no-mètric (NMDS) de les dades amb la funció `metaMDS()` del paquet `vegan` (v.2.6-2)<sup>[86]</sup>, partint de la matriu de dissimilituds de Bray-Curtis computada amb la comanda `vegdist()` del mateix paquet. Les representacions gràfiques derivades ajudaren a examinar l'agrupació adimensional de les mostres entre les espècies de nematode i els entorns de proliferació, i la consegüent similitud entre les comunitats intestinals avaluades.

D'altra banda, per estudiar la dispersió de les mostres es va creure convenient realitzar un *boxplot* o diagrama de caixes múltiple en base als valors de  $\beta$ -diversitat, els quals descriuen la homogeneïtat multivariant de les variàncies entre grups (en aquest cas, les espècies de nematodes en les diferents condicions de creixement). També es va calcular amb la funció `TukeyHSD()` del paquet `stats` (v.4.1.3)<sup>[42]</sup> el test de Tukey necessari per estimar un set d'interval·ls de confiança sobre les diferències entre les mitjanes dels nivells dels factors considerats (de nou, espècie i condició de creixement).

A més a més, es va plantejar la creació d'uns diagrames de barres que permetessin contraposar de forma visual les principals famílies taxonòmiques descrites: 1) al microbioma intestinal de *C. elegans* al llarg del gradient de naturalització, i 2) a la microbiota entèrica de les espècies *C. elegans*, *C. remanei* i *C. briggsae* mantingudes només en entorn natural.

Per últim, es va dissenyar un gràfic de dispersió per reportar el *core microbiome* intestinal de l'espècie *C. elegans* i de *C. remanei*, ambdues crescudes en condicions silvestres. Es van seleccionar els tàxons (en els casos que fou possible fins a nivell de gènere) presents en, com a mínim, el 70% de les mostres naturals avaluades de cada espècie de cuc, i que alhora disposaven d'una abundància relativa mitjana superior al 0.5%. Aquesta abundància relativa promig es defineix com el quocient entre les ocurrencies dels OTUs a cada mostra respecte la riquesa esperada d'espècies a les mateixes mostres (computada amb la funció `specnumber()` del paquet `vegan` (v.2.6-2)<sup>[86]</sup>).

## 4.2 Estudi i modelització computacional dels patrons motors del nematode *C. elegans*

En el transcurs d'aquest segon bloc de treball s'ha adaptat i optimitzat un *pipeline* bioinformàtic d'R apropiat per a la caracterització dels patrons locomotors de les soques N2 i CB4856 de *C. elegans*. Aquest algoritme està constituït per 73 funcions locals, definides explícitament per a segmentar el desplaçament del nematode en dos tipus de moviment, i per estimar els valors parametrizats del model de Langevin associats a la velocitat i la curvatura del cuc al llarg de la trajectòria enregistrada. Cal remarcar que el codi original forma part d'un projecte de tesi encara en desenvolupament de l'estudiant Roger Lloret Cabot, investigador pre-doctoral del CEAB-CSIC.

Val a dir que és impossible procedir amb la gravació de les trajectòries d'alta resolució sense fer una manipulació prèvia dels cucs al laboratori. Resulta necessari, doncs, seguir un rigorós protocol experimental per afavorir el creixement i la proliferació dels nematodes, a fi de disposar d'individus amb l'estadi larvari apropiat per a monitoritzar els seus estats motors. Aquesta secció també ha estat detallada en els mètodes procedimentals dels propers apartats.

Convé destacar que tota la part computacional implementada també s'ha dut a terme de forma local amb el programari R (v.4.1.3) <sup>[42]</sup> i RStudio (v.2022.02.0+443) <sup>[43]</sup>.

### 4.2.1 Mètodes i procediments seguits

#### A. Fase experimental I: preparació de reactius i preservació de *stocks* de nematodes

El primer pas va consistir a elaborar i autoclavar les solucions requerides per a la correcta manipulació dels metazous a nivell experimental, seguint els protocols descrits a l'enciclopèdia de referència *WormBook* <sup>[5]</sup>. D'una banda, es va preparar *S-Buffer* i un compost derivat d'aquest amb un 30% de glicerina (v/v), necessaris per a la correcta preservació criogènica dels *stocks* o vials de nematodes <sup>[87]</sup>.

Paral·lelament, es va produir 1L de *Nematode Growth Medium* (NGM), un medi de cultiu que s'utilitza rutinàriament al laboratori per mantenir individus del gènere *Caenorhabditis* en placa <sup>[87]</sup>. Un dels ingredients d'aquest agar és *KPO<sub>4</sub> buffer*, habitualment conegut amb el nom de tampó fosfat, que també va ser generat a partir de fosfat potàssic dibàsic i tribàsic <sup>[87]</sup>.

Així mateix, es va elaborar tampó M9 i NaOH 8M, dos compostos imprescindibles per portar a terme la sincronització larvària dels cucs: el primer inhibeix l'habilitat ponedora d'ous dels individus adults, i el segon permet dissoldre els organismes indesitjats per preservar únicament els ous en qüestió <sup>[87,88]</sup>.

Amb tot el material necessari a punt, es va procedir a congelar *stocks* de cadascuna de les soques disponibles, N2 i CB4856. Això permetria preservar els individus fins el pas experimental oportú, ja que aquesta espècie té una esperança de vida relativament curta (de 2 a 3 setmanes) <sup>[2]</sup>. Les poblacions de partida es trobaven per separat en dues plaques mare amb agar NGM. Amb aquestes s'inicià el protocol de *freezing* de vials de *C. elegans*, emprant els reactius específics per a aquesta tasca descrits anteriorment i seguint la metodologia indicada al protocol 7 del subcapítol 7. *Freezing and recovery of C. elegans stocks* detallat a l'enciclopèdia *WormBook* <sup>[87]</sup>. Una de les característiques més rellevants d'aquest procediment és que la congelació s'ha de realitzar amb una disminució gradual d'1°C per minut fins els -80°C, d'altra manera els cucs moren ràpidament pel brusc canvi de temperatura <sup>[87]</sup>. A més, el procés només té cabuda quan els nematodes han exhaurit completament l'aliment, en aquest cas *E. coli* OP50 <sup>[87]</sup>.

#### B. Fase experimental II: pre-cultiu d'*E. coli* i preparació de plaques NGM

*C. elegans* s'acostuma a fer créixer monoxènicament al laboratori emprant la soca OP50 d'*E. coli* com a principal font nutricional <sup>[87]</sup>. Aquest bacteri és auxòtrof d'uracil i per tant mostra un creixement

limitat a les plaques de medi NGM <sup>[87]</sup>. Això resulta útil perquè no dificulta la visualització ni el desenvolupament dels nematodes, i alhora permet que aquests s'aparellin millor i s'alimentin contínuament <sup>[87]</sup>. Del contrari, un sobre-creixement del bacteri podria provocar la mort dels cucs <sup>[87]</sup>.

Dit això, es varen aïllar colònies individuals d'*E. coli* OP50 a partir d'un cultiu inicial en placa, les quals foren inoculades posteriorment en un medi ric com *Lysogeny broth* (LB, tradicionalment conegut com Luria-Bertani) <sup>[87]</sup>. Aquest inòcul s'incubà a 21°C *overnight* amb agitació. Passat aquest temps, es varen repartir 0.8mL d'aquest pre-cultiu a cadascuna de les plaques ja preparades amb agar NGM solidificat <sup>[87]</sup>. El volum afegit es va estendre per tota la superfície evitant a tota costa els marges de la pròpia placa Petri, ja que interessava tenir un *patch* d'aliment bacterià disposat cèntricament. Finalment, aquestes plaques NGM amb substrat foren incubades *overnight* a temperatura ambient per desenvolupar la gespa microbiana que serviria d'aliment pels futurs nematodes.

### C. Fase experimental III: descongelació de vials i sincronització de nematodes

Arribats a aquest punt es varen descongelar a temperatura ambient dos vials de nematodes (un de cada soca), el contingut dels quals s'abocà sobre dues noves plaques NGM amb gespa d'*E. coli* OP50 <sup>[87]</sup>. La deposició dels cucs es va realitzar a la zona marginal, no sobre el *patch* directament, per tal d'estimular la cerca activa d'aliment al centre de la placa. Tot seguit, la incubació dels cucs es portà a terme durant 16h, a 21°C i 80% d'humitat <sup>[87]</sup>.

Donat que el cicle vital de *C. elegans* es completa en 3 dies, passat aquest període de temps es recomana fer una transferència d'individus a noves plaques d'NGM amb substrat, especialment quan l'aliment comença a escassejar <sup>[87]</sup>. Així doncs, aquest procés de canvi de placa es va repetir diversos cops fins tenir noves poblacions actives per ambdues soques d'estudi.

Ara bé, tenint en compte que les trajectòries s'havien de generar amb cucs en estadi larvari L4, fou necessari aplicar un protocol de sincronització per treballar amb individus en la mateixa etapa del cicle vital. El procediment en qüestió es troba detallat al subcapítol V. *Preparation of Synchronized Cultures and Specific Stages* del capítol 1 del llibre *Caenorhabditis elegans: Modern Biological Analysis of an Organism* <sup>[89]</sup>. El principi del mètode recau en que l'addició combinada d'hipoclorit sòdic i NaOH elimina tots els nematodes excepte els ous, els quals disposen d'una cutícula protectora que els fa insensibles a aquest tractament <sup>[89,90]</sup>. A més, com que aquesta etapa química s'implementa una vegada s'ha privat amb el tampó M9 la capacitat per pondre més embrions, els que romanen a l'interior dels cadàvers adults s'alliberen en dissoldre l'estructura corporal del progenitor <sup>[89]</sup>.

La resuspensió i posterior incubació dels ous preservats en noves plaques d'NGM amb aliment va permetre disposar de tots els nematodes en la mateixa fase larvària L4 passades 39h exactes <sup>[89,90]</sup>. Aquesta etapa del cicle vital és molt activa des d'un punt de vista motor, i alhora té una mida considerable com per ésser captada a través del sensor de moviment.

### D. Generació de trajectòries individuals d'alta resolució

La següent etapa va consistir a obtenir 10 trajectòries d'alta resolució (a una freqüència de 20 Hz) mitjançant un sistema de seguiment no invasiu de 90 minuts de gravació semi-automàtica. Es seleccionaren aleatòriament 5 individus de cada soca en fase larvària L4, amb l'objectiu d'enregistrar la seva locomoció particular. Aquests es depositaren d'un en un al centre de noves plaques d'NGM sense aliment, disposades sota una estructura il·luminada de càmeres i sensors. El mecanisme de *tracking* permetia gravar el recorregut de cada cuc a partir d'un controlador personalitzat i motoritzat que seguia la posició del seu centre de massa, generant com a resultat uns fitxers d'anotacions que

incloïen les marques temporals i les coordenades espacials X/Y associades al seu moviment, entre d'altres paràmetres <sup>[56]</sup>.

Aquestes trajectòries es processaren amb Eclipse, un IDE de Java que permet dissenyar aplicacions en altres llenguatges de programació com Python o Perl <sup>[91]</sup>. En aquest cas, el programa implementat amb Eclipse es trobava ja disponible al laboratori, i es va usar per parametritzar les opcions d'imatge, ajustar i calibrar la configuració de les càmeres en el procés de *tracking*, mesurar en temps real els termes motrius derivats de les imatges capturades i reconstruir les trajectòries globals dels nematodes. Així doncs, els arxius de desplaçament generats es varen guardar en format .CSV per ser analitzats amb el posterior algoritme computacional, atès que contenien les dades de comportament quantificat necessàries.

### E. Simplificació, optimització i implementació d'un *pipeline* bioinformàtic per a la caracterització de trajectòries

L'última activitat del pla de treball es va centrar en modelitzar el patró de comportament dels 10 individus considerats al pas anterior. Per fer-ho, es va emprar un algoritme d'anàlisi de trajectòries implementat en llenguatge R i constituït per 13 fitxers de codi interdependents. Tots aquests *scripts* foren proporcionats per Roger Lloret, un estudiant pre-doctorant del centre d'investigacions CEAB-CSIC que havia dissenyat part d'aquest procés computacional en el seu projecte de tesi (encara no finalitzat).

La tasca desenvolupada en el present treball va consistir a adequar i optimitzar el *pipeline* disponible amb l'objectiu d'examinar posteriorment les trajectòries enregistrades amb el sistema de *tracking*. La finalitat de tota aquesta investigació era contrastar la motilitat de les diferents soques, d'una banda segmentant en dos tipus de moviment el patró de locomoció seguit pels nematodes, i d'una altra estimant els valors cinemàtics parametritzats del model de Langevin <sup>[44]</sup>.

La revisió dels arxius del *pipeline* i la introducció dels canvis pertinents va resultar en un conjunt de 73 funcions definides en l'entorn local d'execució. D'aquestes, només 11 s'empren de forma directa, ja que la resta constitueixen sub-funcions que s'inicialitzen en aplicar les parentals. L'organització dels diversos fitxers de codi és la següent:

1. `data_processing.R`: en aquest arxiu es varen definir funcions adequades per a la importació de les trajectòries obtingudes amb el procediment de *tracking*, com per exemple `read_celegans()`.
2. `motion_magnitudes.R`: en aquest fitxer es varen establir comandes útils per a l'estudi d'algunes magnituds cinemàtiques, com l'angle de gir entre punts consecutius de la trajectòria (funció `turnangl()`) o la distància euclidiana entre els mateixos (funció `eudist()`).
3. `distributions_functions.R`: aquest *script* estava dedicat a la caracterització de la distribució de probabilitat mixta gamma i normal (específicament la funció de densitat de probabilitat, la funció de la seva distribució acumulada, la funció de quantils i la funció de generació de valors aleatoris).
4. `crude_parameter_estimation.R`: les comandes definides en aquest document permetien fer una estimació en brut dels paràmetres de la part elemental gamma de la distribució mixta (com serien *shape* i *rate*, computades a partir de les funcions `mll_grate/gshape()` respectivament).
5. `speed_distribution_fit.R`: les funcions descrites aquí resultaren útils per ajustar i modelar la velocitat de desplaçament dels individus. Destaquen les comandes `save_pop_smod()` i



`fit_speed_dgammanorm()` (útils per ajustar una distribució combinada gamma + normal de la velocitat dels nematodes), així com `backtr_ldgn_mod()` (adequada per descriure la parametrització lineal de la distribució mixta computada amb les altres funcions).

6. `common_tools_xtip_classification.R`: en aquest arxiu es varen concretar les funcions necessàries per definir els “*tips*” (com `get_tseries_tips()` o `lab_true_reortips()`) i els dos estats motors de referència a considerar en la classificació del moviment: el serpenteig en sentit lineal o *crawling*, i la reorientació o *looping* (que inclou inversions i girs).

Els *tips* són vectors d’observacions temporals consecutives amb el mateix signe en el seu diferencial, i s’utilitzen per assignar les velocitats d’aquell espai temporal a la part elemental corresponent de la distribució probabilística mixta, amb l’objectiu de diferenciar els dos tipus de moviment mencionats. Els canvis d’estat motor només poden ocórrer quan es produeix un canvi de signe en el diferencial esmentat. Es varen anomenar així pel fet de tenir una forma gràfica de punta (*tip* en anglès), i perquè aquests canvis d’estat venien associats de manera implícita a un canvi en la velocitat amb què es movia el nematode.

7. `tips_xhist_rootSSQ_class.R`: les comandes d’aquest *script* ajudaren a factoritzar la trajectòria dels individus en els dos tipus de moviment mencionats. És a dir, varen permetre classificar els *tips* en funció de la distribució probabilística més idònia. Cal destacar la funció parental `save_pop_etho()`, la qual és capaç d’inicialitzar fins a 7 subnivells d’execucions amb una cinquantena de funcions *helper*. Aquesta s’encarrega de guardar, per a tots els individus, les taules de tipus etograma amb els estats motors de cada fase de la trajectòria <sup>[92]</sup>.
8. `empirical_distributions.R`: aquest fitxer de codi recollia funcions adients per al càlcul de les distribucions empíriques de les dades, com `get_pdf_fdtab()` (amb la qual es varen generar taules de freqüències útils per a comparar aquestes densitats amb les teòriques o estimades).
9. `behavioural_annotation.R`: en aquest arxiu es varen definir funcions per a l’estudi numèric i gràfic dels dos estats motors avaluats. La funció `compact_begend_states()` permetia compactar les coordenades inicials i finals de cada estat motor, mentre que les funcions `import_idetho()` o `plot_ethogram()` servien per importar i representar les taules d’etograma amb els models de velocitat computats.
10. `distribution_fits.R`: les dues comandes descrites en aquest fitxer foren útils per ajustar les distribucions probabilístiques associades als diferents còmputs.
11. `plot_functions.R`: aquí s’inclogueren funcions suplementàries de representació visual de les dades (com `plot_layout_fdtab()`, que servia per il·lustrar la taula de densitat de freqüències calculada en els arxius anteriors).
12. `dphi_distribution_fit.R`: les funcions locals presentades en aquest arxiu es varen preparar per ajustar el model de velocitat angular dels individus al llarg de la trajectòria. Les comandes `save_pop_dphimod()` i `fit_dphi_dcauchy()` varen possibilitar l’ajust i la descàrrega del model mencionat.
13. `rc_save_speed_analysis.R`: aquest és l’*script* principal des d’on es va realitzar l’anàlisi complet de les trajectòries i a partir del qual s’activaren i executaren les funcions dels anteriors fitxers de codi.

Dit això, l'aplicació conjunta de tots aquests *scripts* va permetre caracteritzar la motilitat de *C. elegans* i la seva variabilitat entre individus d'una mateixa soca i també entre soques. A continuació es descriuen les etapes vinculades a la caracterització del comportament motor dels nematodes, realitzades per duplicat donat que es treballava amb dues soques diferents.

### **Càrrega de les trajectòries d'alta resolució a R**

La implementació de l'algoritme requeria inicialment apujar els arxius de trajectòries en format .CSV, mitjançant la comanda `read_celegans()` definida al fitxer #1. D'aquesta manera, es varen contenir en objectes de R els valors associats a les marques temporals, les coordenades espacials i les velocitats dels 5 nematodes d'una mateixa soca, cadascun dels quals estava assignat a un codi identificador propi i intransferible.

### **Neteja de les trajectòries i còmput de magnituds cinemàtiques**

Seguidament, les dades carregades foren depurades a fi de suprimir moments de pausa en què el nematode romania quiet. És a dir, es varen eliminar aquells instants en els quals la velocitat deueia per sota d' $1 \cdot 10^{-14}$  mm/s, l'angle de gir era 0 i la distància euclidiana entre punts resultava menor a  $1 \cdot 10^{-10}$  mm.

D'altra banda, també es varen inferir algunes magnituds cinemàtiques d'interès, com  $dt$  (el diferencial de temps entre cada parella consecutiva de valors de la trajectòria),  $dx$  (la distància euclidiana entre els punts consecutius) i  $dphi$  (l'angle de gir entre coordenades X/Y també de punts consecutius). Per fer-ho s'utilitzaren funcions com `turnangl()` i `eudist()`, definides a l'*script* #2.

A més a més, amb les coordenades espacials de cada punt es representaren amb un diagrama de línia les trajectòries seguides per cadascun dels individus.

### **Generació dels models de velocitat**

Arribats a aquest punt, s'inicialitzà la funció parental `save_pop_smod()` (arxiu #5) per computar, ajustar i guardar localment els models de velocitat de tots els membres de la soca considerada. El model de cada cuc en particular es va importar posteriorment a R amb la comanda `import_idsmod()` del mateix *script*. Finalment, amb la funció `backtr_ldgn_mod()` es realitzà una conversió a escala lineal dels valors calculats, amb l'objectiu de fer-los més fàcilment interpretables per a l'usuari donat que es trobaven inicialment en escala logarítmica.

### **Diferenciació dels estats motors i creació de l'etograma**

La següent etapa del processat va servir per absorbir o assignar els *tips* a una de les dues condicions motores. Això s'assolí amb la comanda `save_pop_etho()` del fitxer de codi #7, a partir de la qual fou possible activar tot un seguit de funcions complementàries. Com a resultat, s'originaren taules de tipus etograma que classificaren cada fase de la trajectòria a un estat de locomoció concret. Un etograma és un inventari de conductes de moviment animal objectives i estereotipades, emprades típicament per a comprendre la prevalença de cada tipus de comportament exhibit <sup>[92]</sup>. Aquestes taules solen informar de la durada i les posicions relatives de cada interval de la trajectòria associat a un determinat patró de desplaçament <sup>[92]</sup>.

Al llarg de l'execució es varen anar actualitzant alhora les taules de freqüències que descriuen la distribució probabilística mixta del comportament motor, així com l'error associat a una classificació equívoca del *tip* en una distribució inadequada.

Tal i com s'ha comentat abans, els *tips* són vectors que, de forma planera, representen cada estat motriu a partir d'un segment temporal de la trajectòria que té un inici i un final. L'algoritme de designació que es va definir en aquesta part calculava de manera iterativa la versemblança que tenia cada *tip* de formar part o bé de la distribució gamma o bé de la gaussiana. Per tant, els *tips* es varen anar absorbint de manera incremental per anar omplint les distribucions estadístiques elementals, que alhora descriuen una fase de moviment específica.

Les taules d'etograma generades es varen representar mitjançant un diagrama de barres per tenir una percepció global de les diferents fases motores al llarg del temps de trajectòria, utilitzant la comanda `plot_ethogram()` definida en el fitxer de codi #9. Un gràfic derivat d'aquest va resultar útil per contemplar simultàniament els canvis de velocitat del nematode respecte el temps de locomoció.

### **Ajust dels models de velocitat angular**

Paral·lelament, amb la funció `save_pop_dphimod()` (detallada a l'*script* #12) fou possible computar i guardar els models de velocitat angular per a tots els individus d'una mateixa soca. La velocitat angular descriu un concepte físic que permet estudiar el moviment de rotació periòdic d'un ésser o un objecte a partir de l'angle girat per unitat de temps, i s'acostuma a presentar en radians per segon (rad/s) <sup>[93]</sup>. Finalment, els models mencionats es varen ajustar a una distribució de probabilitat contínua Cauchy-Lorentz amb la funció `fit_dphi_dcauchy()` (definida també a l'*script* #12). Aquesta distribució probabilística és destacable en el camp de l'espectroscòpia i l'estudi físic dels fenòmens de ressonància, tot i que aquí es va emprar amb finalitats més biològiques per estudiar el moviment ondulatori dels cucs <sup>[94]</sup>.

Els valors ajustats de velocitat angular per a cadascun dels individus foren representats en un histograma amb la distribució teòrica de les dades i diferenciant els dos estats motors possibles.

### **Parametrització del model de Langevin**

El model de Langevin és una aproximació físic-matemàtica desenvolupada per l'autor del mateix nom que s'utilitza habitualment per estudiar fenòmens de dinàmica molecular <sup>[95]</sup>. Destaca per tenir aplicabilitat en els processos de difusió, oscil·lació harmònica i desplaçament Brownià, ja que permet descriure l'evolució d'un sistema de variables sotmès a una combinació de forces fluctuants (és a dir, aleatòries) <sup>[37,95,96]</sup>.

Els fonaments matemàtics del model en qüestió han estat adaptats per al seu ús en Ecologia i altres disciplines relacionades, en especial per discernir els mecanismes subjacents en les estratègies de locomoció d'animals terrestres. Aquest model també resulta propici per caracteritzar l'equilibri entre fases d'exploració i explotació de l'entorn, que combinades fan més eficient la cerca d'aliment <sup>[29,37]</sup>. Les equacions diferencials estocàstiques utilitzades en aquest *framework* són les següents <sup>[97]</sup>:

$$(Eq. 1) \quad \frac{dv}{dt} = -\gamma(v - v_s) + \sigma_n \quad (Eq. 2) \quad \frac{d\phi}{dt} = \frac{\beta + \sigma_\phi}{v}$$

L'equació 1 descriu la variació de velocitat de l'individu al llarg del temps de trajectòria, on  $v$  fa al·lusió a la velocitat d'un organisme en un punt concret de la trajectòria i  $v_s$  a la velocitat mitjana considerant el recorregut complet <sup>[97]</sup>. El paràmetre  $\sigma_n$  fa referència a la desviació típica associada a la mesura de la velocitat mitjana <sup>[97]</sup>. El terme  $\gamma$  descriu la intensitat de la força de propulsió, o el que és el mateix, la velocitat de persistència <sup>[97]</sup>. Es tracta d'un paràmetre macroscòpic de validació del model que mesura

quant temps la velocitat s'ha mantingut per sobre del valor promig <sup>[97]</sup>, però no ha estat avaluat en el present Treball de Fi de Màster donada la complexitat existent en la seva estimació.

D'altra banda, l'equació 2 descriu el canvi en la curvatura de l'organisme al llarg del temps de trajectòria <sup>[97,98]</sup>. El terme  $\beta$  és una mesura de la força de torsió que pren l'individu mentre avança, i es computa a partir del producte de velocitat mitjana per velocitat angular mitjana <sup>[97]</sup>. Com abans,  $v$  fa al·lusió a la velocitat d'un organisme en un punt concret de la trajectòria, i  $\sigma_\phi$  avalua la desviació típica associada a la mesura de l'angle de gir entre punts consecutius del recorregut <sup>[97]</sup>.

Dit això, tots aquests paràmetres foren calculats per als 10 individus analitzats a partir dels coeficients ajustats dels models de velocitat i velocitat angular computats anteriorment.

## 5 Resultats

### 5.1 Descripció del perfil microbià intestinal del gènere *Caenorhabditis*

La consecució del primer bloc d'estudi descrit en el present treball ha culminat amb la generació de diversos gràfics que il·lustren la variació de la microbiota entèrica dels membres del gènere *Caenorhabditis* entre les diferents condicions ambientals valorades al meta-anàlisi, tal i com es descriu al subapartat H de la secció 4.1.1.

La figura 3 correspon a una representació NMDS de les mostres de microbioma finalment processades, les quals queden identificades per color en funció de la condició de creixement de l'hoste de procedència. Aquestes categories de desenvolupament constitueixen un gradient de naturalització, que va des d'una proliferació assilvestrada (designada a la llegenda com *Natural*) fins a un manteniment en condicions semi-artificials al laboratori (etiquetat com *Experimental*), passant pel desenvolupament en microcosmos (anomenat *Microcosms*) i l'enriquiment amb *E. coli* OP50 després de l'aïllament (etiquetat com *Lab enriched*).

Es va decidir preservar com a categories independents les condicions de creixement homòlogues tractades en *datasets* diferents, amb l'objectiu de distingir possibles diferències en l'agrupació de les mostres de cada metagenoma. Per aquesta raó, s'infereixen tres categories experimentals, dues CeMbio (una per cada *dataset*) i una més genèrica que recull principalment les lectures de *Dirksen\_experimental* (que treballava amb una microbiota model diferent a la de CeMbio). Quelcom similar succeeix amb les categories d'enriquiment al laboratori, diferenciades segons els estudis d'origen tot i respondre a una mateixa condició de proliferació.

L'escalat multidimensional no-mètric implementat, que cerca solucions estables a partir d'inicialitzacions aleatòries de les dades, resulta útil per inferir el nivell de similitud i la corresponent agrupació de les mostres avaluades. El còmput en qüestió ha condensat la informació procedent de dades multidimensionals en un diagrama 2D, permetent així aproximar la posició original d'una comunitat a l'espai  $k$ -dimensional de la forma més acurada possible. Com més properes es troben les mostres en l'espai d'ordenació, més similars esdevenen les comunitats microbianes associades a aquestes entre sí.

Les formes el·lipsoïdals permeten tenir una percepció de l'àrea ocupada per la massa principal de punts (o mostres del meta-anàlisi en aquest cas). Com més amplis són els el·lipsoïdes, més dispersió hi ha entre les mostres d'una mateixa condició. Aquestes àrees han estat computades de manera que englobin el 95% de les dades de la categoria corresponent.

Una bona manera de comprovar si aquesta ordenació resumeix idòniament les distàncies observades entre les mostres és calculant el valor d'*stress* associat, que en aquest cas era de 0.213 (un resultat prou òptim).

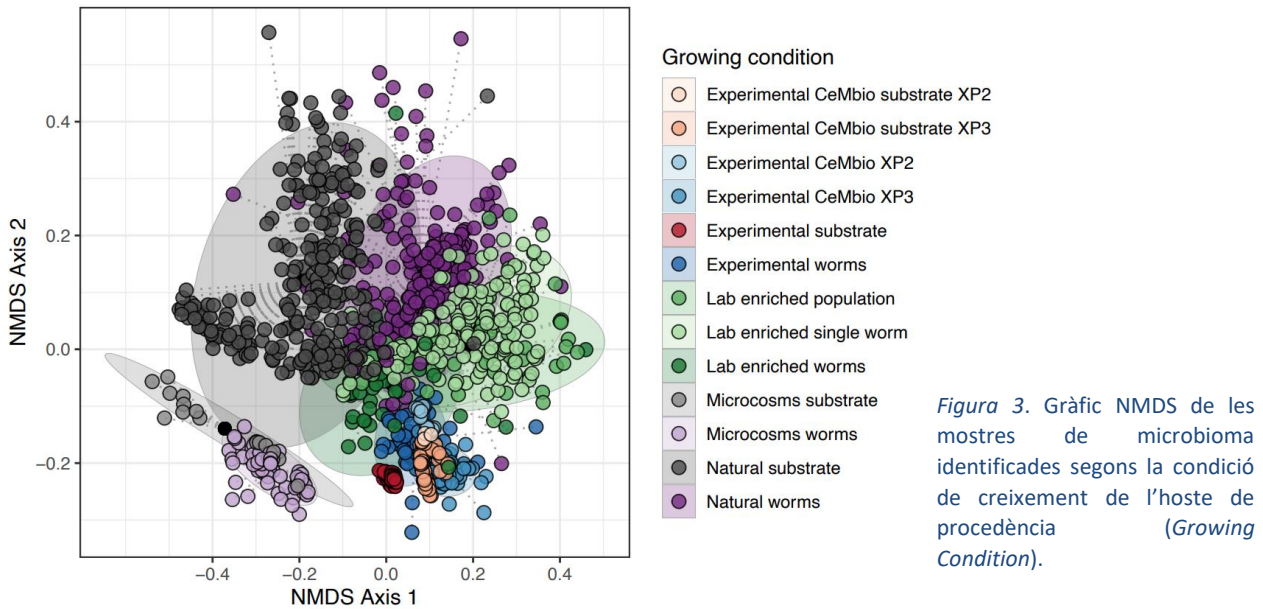


Figura 3. Gràfic NMDS de les mostres de microbioma identificades segons la condició de creixement de l'hoste de procedència (*Growing Condition*).

Tal i com es pot veure a la figura 3, els substrats poden tenir una rellevància considerable en la constitució de la microbiota intestinal dels nematodes, siguin quines siguin les seves condicions de creixement. Per aquesta raó, es va creure convenient suprimir totes les mostres de sòl de l'*input*, amb l'objectiu d'estudiar la dispersió de la comunitat intestinal amb independència dels substrats. Això es pot observar a la figura 4. En aquesta il·lustració, a més a més, les condicions de creixement equivalents foren finalment condensades en una sola categoria per reduir la complexitat d'interpretació, tot i que es mantingueren encara separades les situacions de proliferació semi-artificial pel fet de treballar amb microbiotes model força diferents.

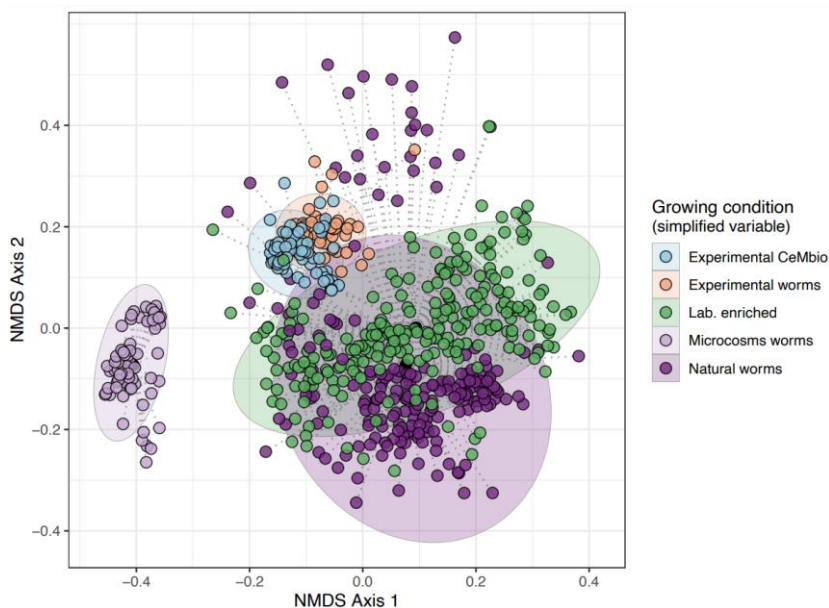


Figura 4. Gràfic NMDS de les mostres de microbioma identificades segons la condició de creixement de l'hoste de procedència (*Growing Condition*), després de resumir les categories i eliminar les mostres de microbioma de substrat.

L'últim escalat NMDS computat i graficat (veure figura 5) va servir per concatenar les condicions de creixement simplifiades de la figura 4 amb les diferents espècies de *Caenorhabditis* considerades en el meta-anàlisi. Cal remarcar, però, que en els estudis originals no es contemplaren totes les combinacions d'espècie i entorn de proliferació, d'aquí que a la figura només apareguin aquelles disponibles a les dades originals.

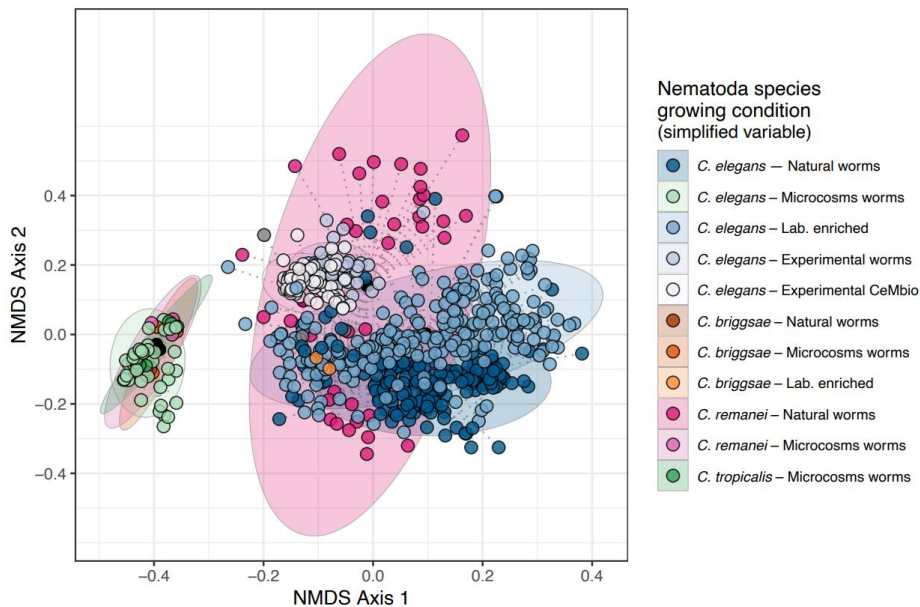


Figura 5. Gràfic NMDS de les mostres de microbioma identificades segons l'hoste de procedència i la seva condició de creixement, després de resumir les categories i eliminar les mostres de microbioma del substrat.

La figura 6A està constituïda per un diagrama de caixes múltiple a partir del qual és possible estudiar la dispersió de les mostres segons els valors de  $\beta$ -diversitat computats. Aquests *boxplots* són formes estandarditzades de mostrar la distribució de les dades en base a un resum de 5 valors (mínim, primer quartil (Q1), mediana, tercer quartil (Q3) i màxim); a més, permeten indicar si les dades són simètriques, quina és la seva agrupació i si estan esbiaixades i de quina forma. L'eix d'abscisses configura la distància de cada punt al centre de l'el·lipse que els conté en el traçat NMDS, mentre que a l'eix d'ordenades s'indiquen les diferents combinacions d'espècie de metazou i condició de proliferació avaluades anteriorment.

D'altra banda, la figura 6B recull part del resultat del test de Tukey que mostra, en forma d'interval de confiança al 95%, una estadística multivariant de comparació de distàncies mitjanes al centre entre les diverses combinacions de grups valorades. En particular, es mostren les columnes *diff* (la diferència entre els diferents valors de distància mitjana al centre dels grups contrastats), *lwr* i *upr* (els límits inferior i superior de l'interval), i *p adj* (el p-valor ajustat).

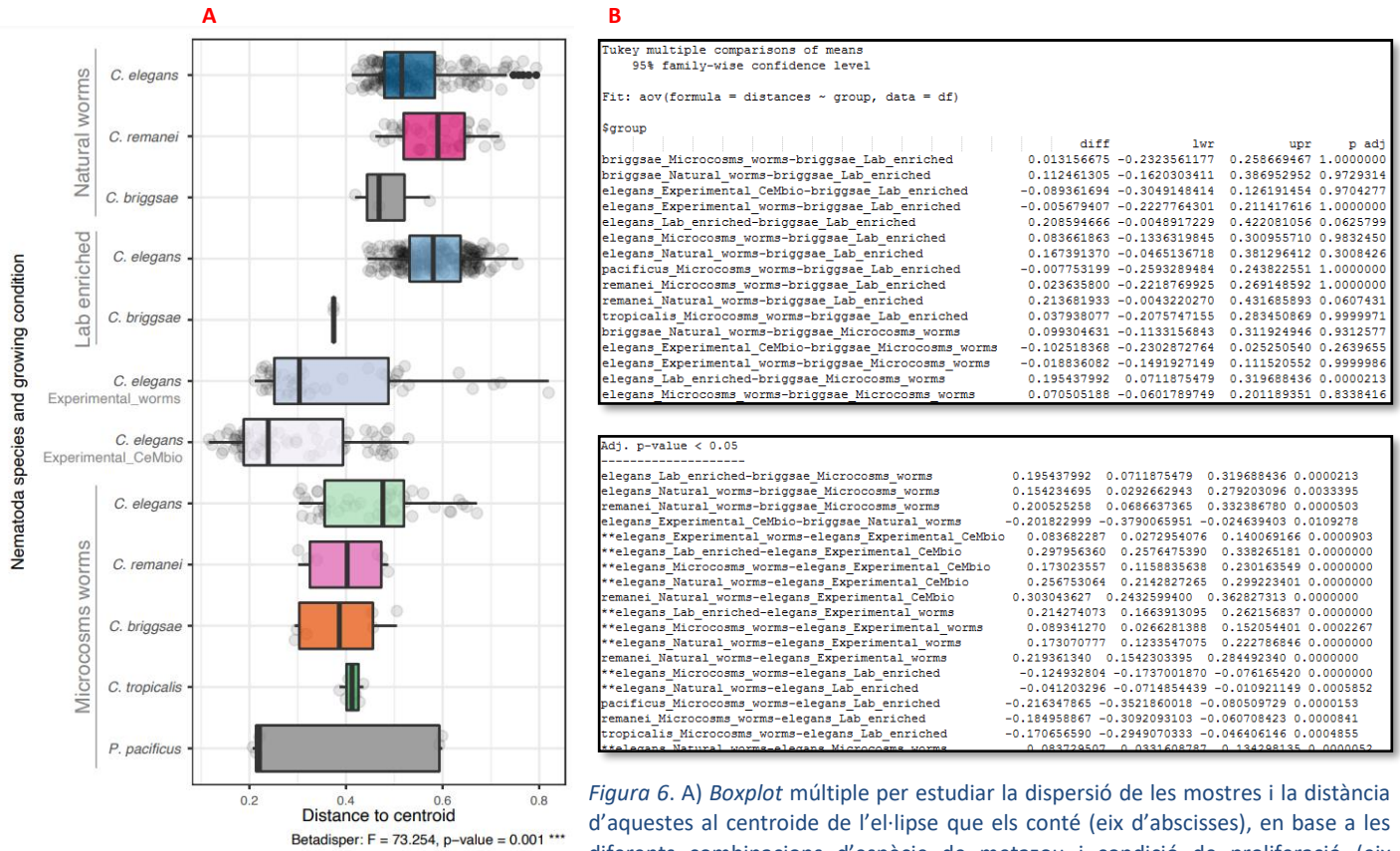


Figura 6. A) Boxplot múltiple per estudiar la dispersió de les mostres i la distància d'aquestes al centroid de l'el·lipse que els conté (eix d'abscisses), en base a les diferents combinacions d'espècie de metazou i condició de proliferació (eix d'ordenades). B) Càlcul resultant del test de Tukey per a la comparació de distàncies mitjanes al centroid entre les diverses combinacions de grups valorades.

Per un altre costat, el gràfic descriptor de la figura 7 ajuda a identificar les principals famílies taxonòmiques descrites als OTUs, amb l'objectiu de contrastar posteriorment les mostres de microbioma dels cucs de l'espècie *C. elegans* al llarg del gradient de naturalització. Els tàxons amb una elevada abundància relativa en una o més condicions de creixement foren enumerats per tal d'identificar-los fàcilment de la resta de membres.

En canvi, a la figura 8 s'estudia la taxonomia bacteriana intestinal (també a nivell de família) de les espècies *C. elegans*, *C. remanei* i *C. briggsae* mantingudes únicament en condicions de creixement natural. Com abans, els tàxons amb major abundància relativa foren etiquetats numèricament per tenir una clara comprensió de la llegenda mostrada.

En ambdós casos, el llindar de representació mínima dels tàxons a les mostres es fixà a 0.5, resultant en un selecció semi-conservadora de les famílies microbianes localment més abundants i amb una ocurrencia intermèdia en els datasets.



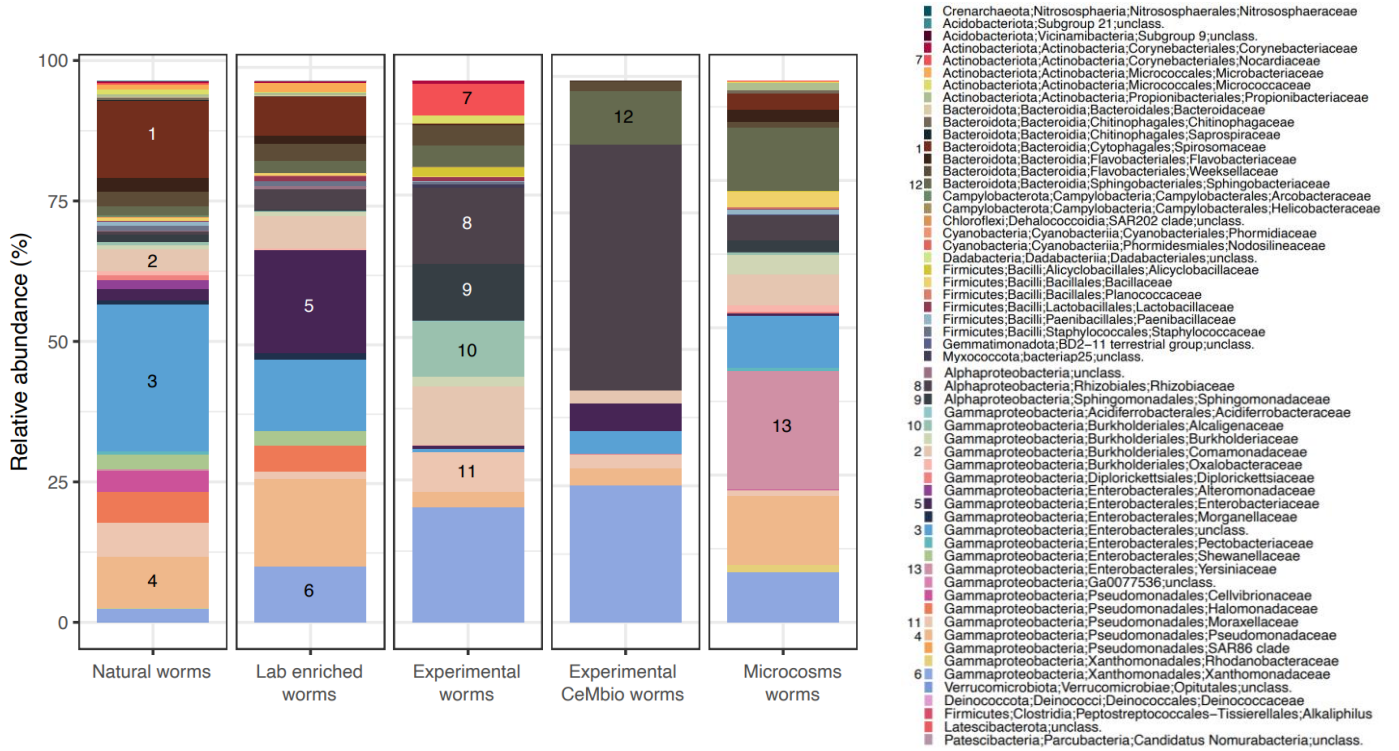


Figura 7. Representació de la comunitat intestinal de *C. elegans* diferenciant les diferents mostres segons l'entorn de creixement de l'individu de procedència, indicant les classificacions taxonòmiques dels OTUs en base al color descrit a la llegenda. A més, es mostren enumerats els tàxons amb una elevada abundància relativa en una o més condicions de creixement.

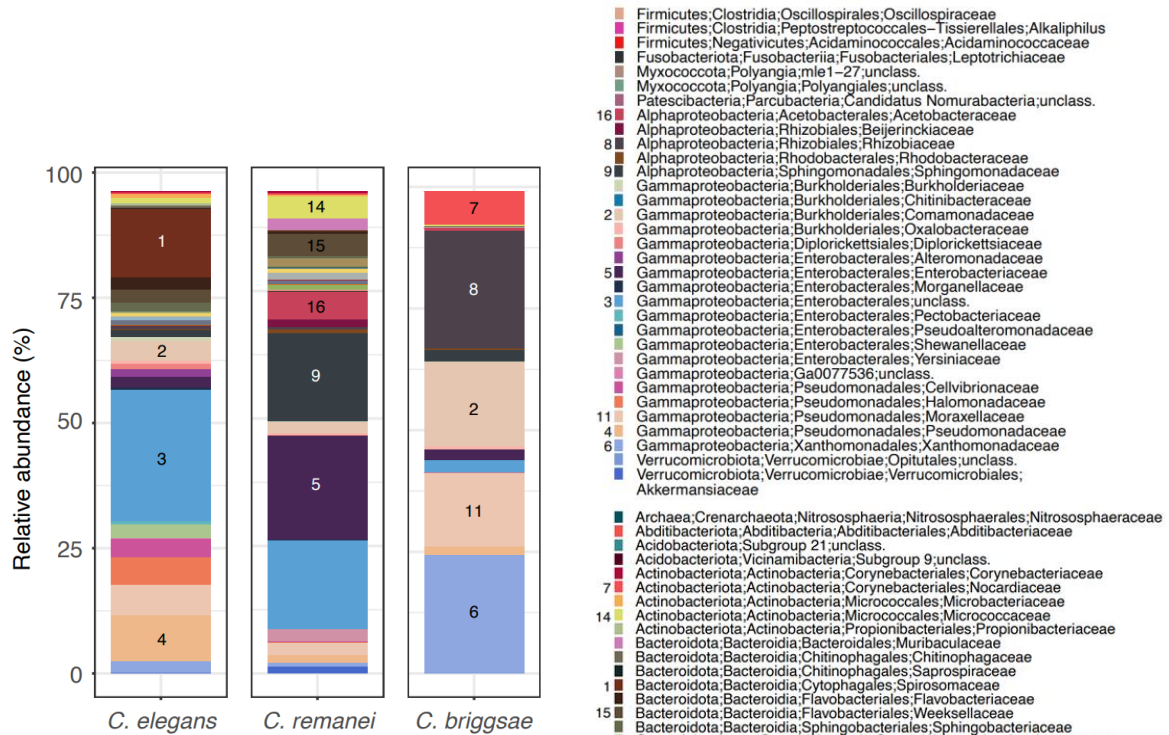


Figura 8. Representació de la comunitat intestinal de *C. elegans*, *C. remanei* i *C. briggsae* aïllats de l'entorn natural, indicant les classificacions taxonòmiques dels OTUs en base al color descrit a la llegenda. A més, es mostren enumerats els tàxons amb una elevada abundància relativa en una o més espècies.



Per últim, les figures 9 i 10 il·lustren a través d'un diagrama de dispersió els tàxons microbians que conformen el *core microbiome* intestinal de *C. elegans* i *C. remanei*, respectivament. Ambdues espècies de nematodes foren valorades únicament en condicions naturals de creixement. A l'eix d'abscisses es representa el nº de mostres avaluades, les quals responen a la combinació d'espècie i categoria de desenvolupament indicada. A l'eix d'ordenades, en canvi, s'expressa l'abundància relativa mitjana dels OTUs en format percentual, derivada del quocient entre l'ocurrència d'aquests a cada mostra respecte la riquesa esperada d'espècies a les mateixes mostres.

Els *core microbiome* foren seleccionats en base a un riquesa superior al 70% (és a dir, cada OTU s'havia de trobar com a mínim en el 70% o més de les mostres considerades) i un llindar d'abundància relativa major a 0.5. Com a resultat, s'identificaren 12 membres en el *core microbiome* de *C. elegans* i 21 membres en el de *C. remanei*; tots aquests foren acolorits després en funció del fílum al qual pertanyen.

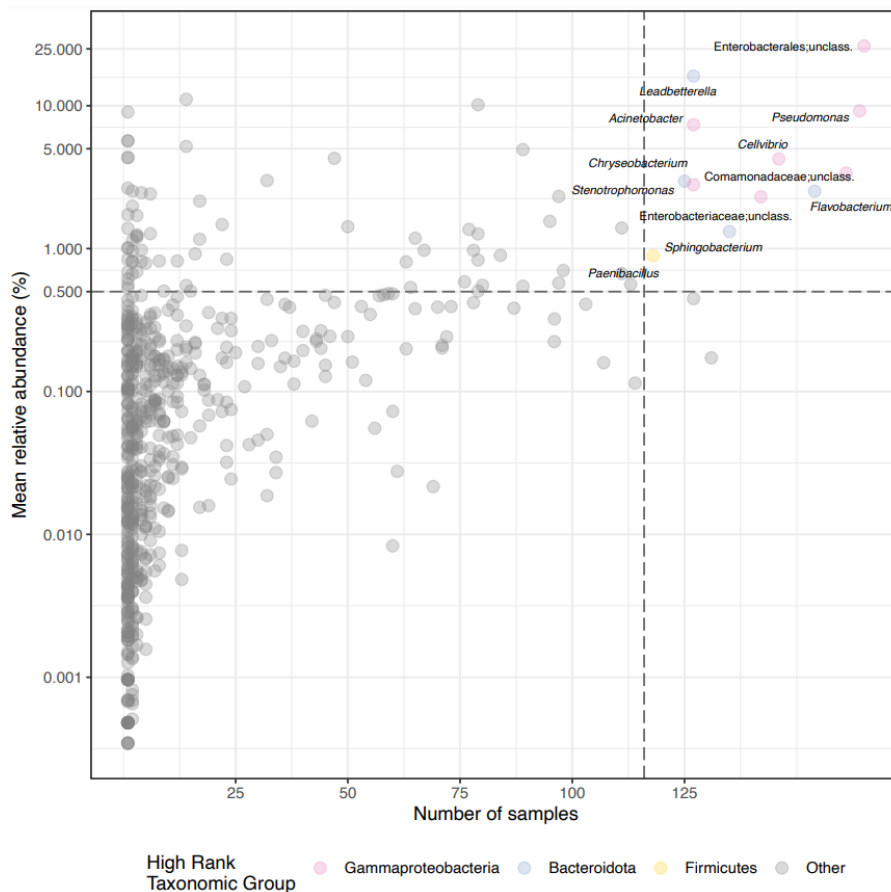


Figura 9. Diagrama de dispersió que il·lustra l'abundància relativa mitjana dels OTUs (eix d'ordenades) representats en base al nº de mostres avaluades (eix d'abscisses) de *C. elegans*. S'indiquen els gèneres bacterians presents en almenys el 70% de les mostres amb una abundància major al 0.5%, indicant la família corresponent en base al color.

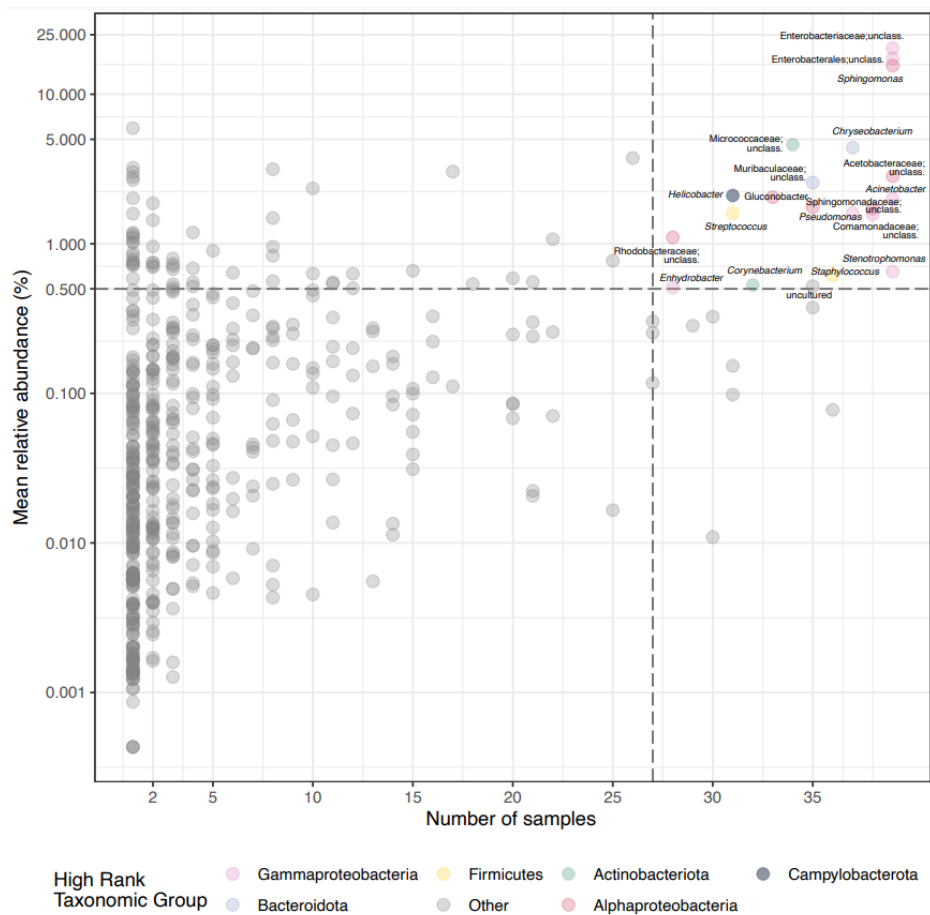


Figura 10. Diagrama de dispersió que il·lustra l'abundància relativa mitjana dels OTUs (eix d'ordenades) representats en base al nº de mostres avaluades (eix d'abscisses) de *C. remanei*. S'indiquen els gèneres bacterians presents en almenys el 70% de les mostres amb una abundància major al 0.5%, indicant la família corresponent en base al color.

### 5.2 Caracterització dels patrons de locomoció de *C. elegans*

La implementació de l'algorisme d'anàlisi de trajectòries, descrit al subapartat E de la secció 4.2.1, ha permès generar els *outputs* necessaris per a contrastar els patrons de moviment de les soques N2 i CB4856 de *C. elegans*. A continuació només es detallen els resultats obtinguts per a un mateix individu de cada soca a mode d'exemple, a fi d'evitar redundància en la presentació dels resultats. La resta de gràfics i còmputos pertanyents als 8 individus restants poden consultar-se a l'Annex II.

La figura 11 està constituïda per dos gràfics de línia que representen el recorregut d'un individu d'una de les dues soques manipulades. Els cucs iniciaren la seva trajectòria al centre de la placa (punt amb coordenades 0,0), i el desplaçament es mesurà en mil·límetres.

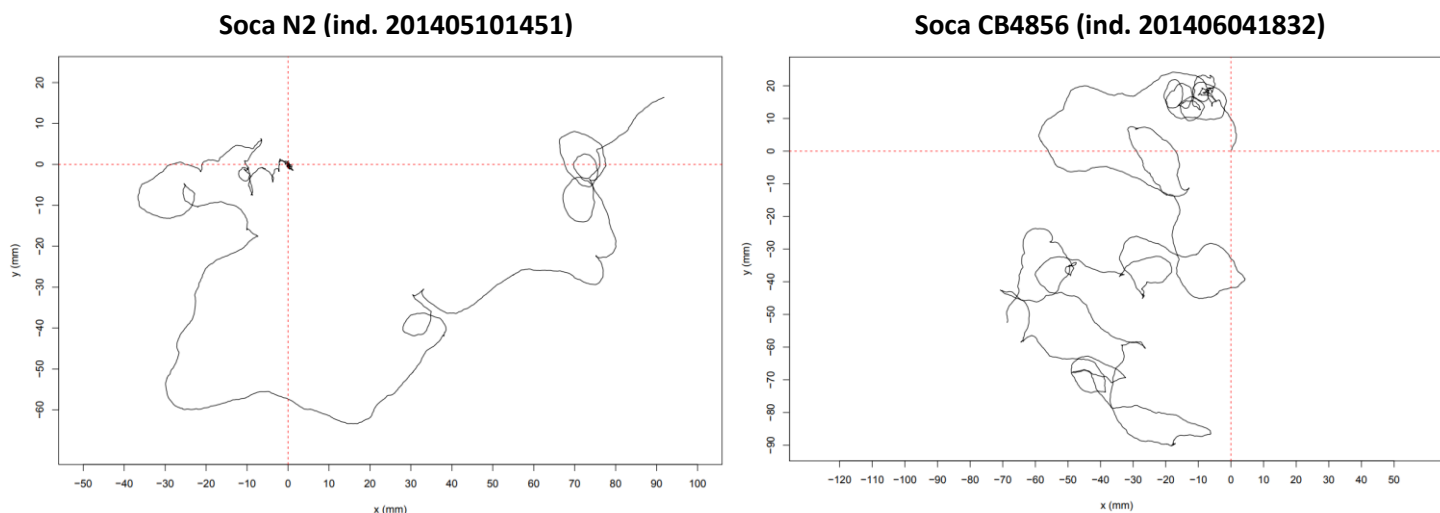
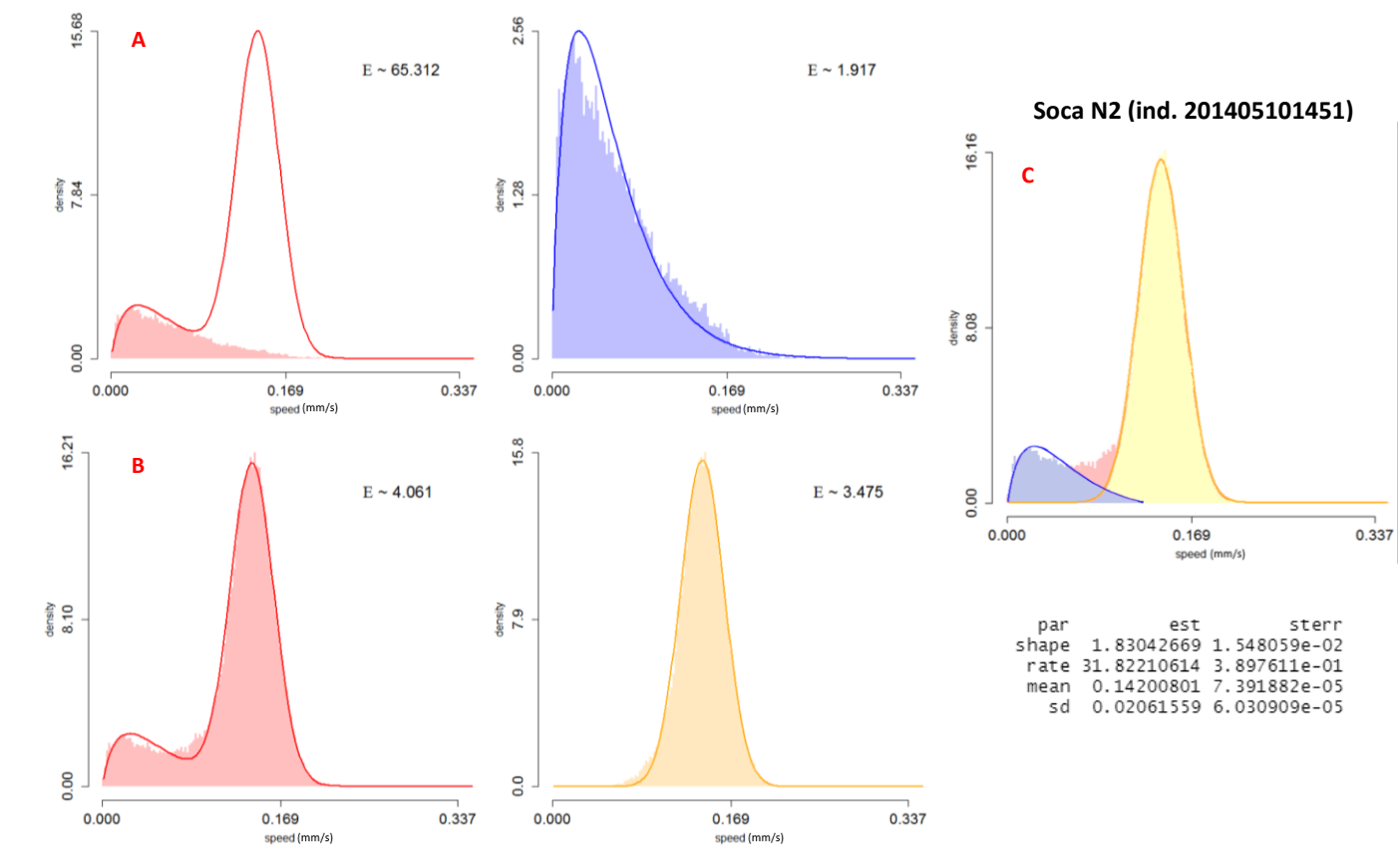


Figura 11. Representació de la trajectòria dels individus 201405101451 (soca N2) i 201406041832 (soca CB4856) de *C. elegans* en base a les coordenades cartesianes X/Y mesurades en mil·límetres.

D'altra banda, les distribucions de densitat de les velocitats lineals seguides per un individu de cada soca poden consultar-se a les figures 12 i 13. El mosaic de quatre gràfics (part esquerra de la imatge, seccions A i B) defineix l'assignació dels segments de trajectòria o *tips* a cadascuna de les distribucions probabilístiques elementals, gamma (secció A) o gaussiana (secció B). Els gràfics acolorits en vermell representen el conjunt d'ambdues en una distribució mixta, mentre que els gràfics acolorits en blau o en groc il·lustren les distribucions elementals mencionades. Al gràfic de la secció C es mostra la mateixa distribució combinada però diferenciant per colors cadascuna de les parts elementals que la constitueixen.

Adicionalment, s'indica l'error associat a una classificació equívoca dels *tips*. El procés comença absorbint aquells que siguin propis d'una distribució de valors gamma, donat que els valors associats a aquesta són clarament distintius. A mesura que el còmput continua i s'omplen les parts elementals de la distribució mixta, l'error disminueix fins a estabilitzar-se. El programa és capaç d'avaluar si l'assignació d'un *tip* a una distribució elemental concreta implica un augment o una reducció d'aquest l'error de càlcul; si es produeix un increment, aleshores es re-assigna el *tip* a l'altra part de la distribució mixta.



**Figura 12.** Distribucions de densitat de les velocitats lineals seguides per l'individu 201405101451 de la soca N2 de *C. elegans*. A) Representació de la distribució mixta (en vermell) i individual (en blau) dels segments de trajectòria assignats a la distribució probabilística gamma. B) Representació de la distribució mixta (en vermell) i individual (en groc) dels segments de trajectòria assignats a la distribució probabilística gaussiana. C) Il·lustració conjunta de la distribució combinada diferenciant per colors cadascuna de les parts elementals que la constitueixen. S'indiquen addicionalment els paràmetres que defineixen la distribució gamma (*shape* i *rate*) i normal (*mean* i *sd*).

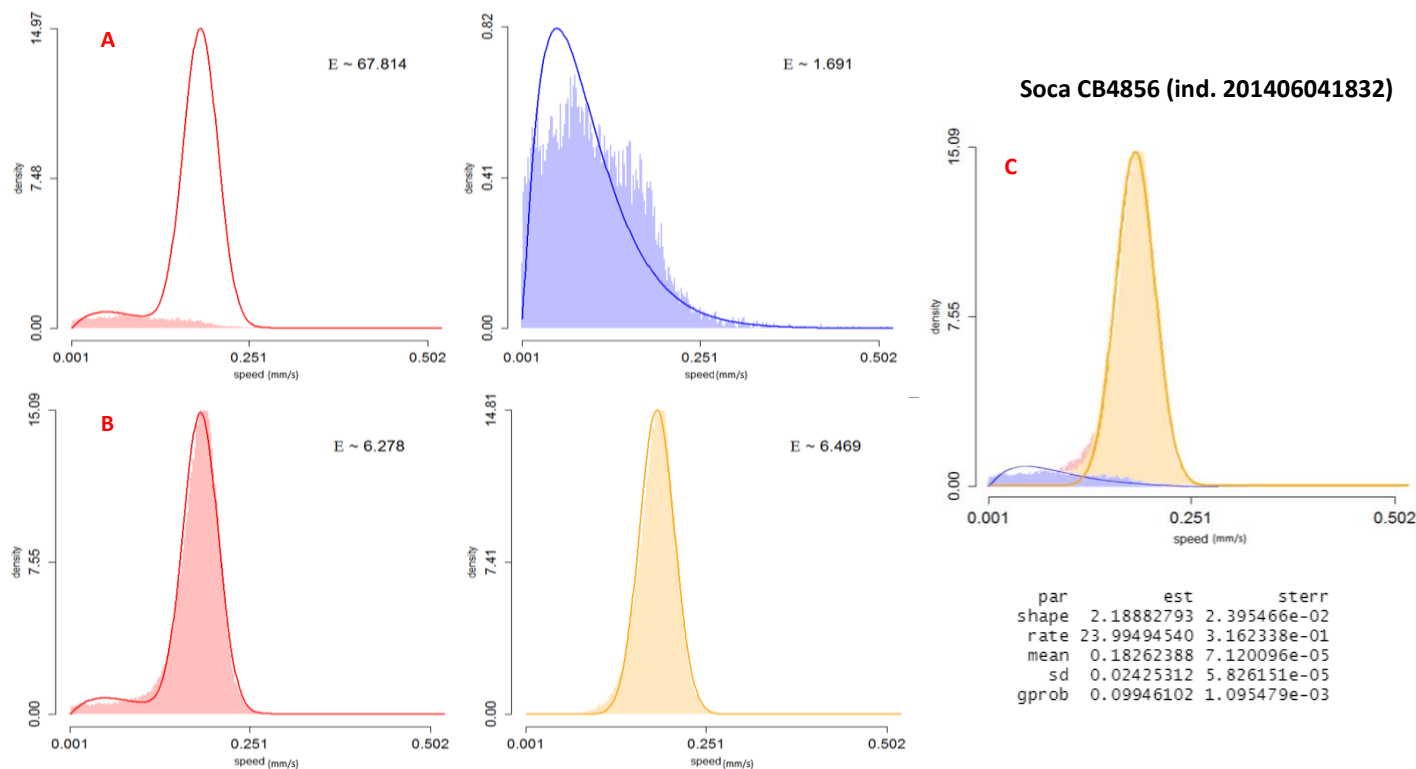


Figura 13. Distribucions de densitat de les velocitats lineals seguides per l'individu 201406041832 de la soca CB4856 de *C. elegans*. A) Representació de la distribució mixta (en vermell) i individual (en blau) dels segments de trajectòria assignats a la distribució probabilística gamma. B) Representació de la distribució mixta (en vermell) i individual (en groc) dels segments de trajectòria assignats a la distribució probabilística gaussiana. C) Il·lustració conjunta de la distribució combinada diferenciant per colors cadascuna de les parts elementals que la constitueixen. S'indiquen addicionalment els paràmetres que defineixen la distribució gamma (*shape* i *rate*) i normal (*mean* i *sd*).

Les figures 14 i 15 combinen els diagrames d'etograma (secció A), derivats de traçar amb diferents colors els dos estats motors contemplats en la taula del mateix nom i al llarg de la trajectòria, amb gràfics de barres que aporten informació relativa a la velocitat del nematode en cada tram del recorregut (secció B). Les zones blaves defineixen estats de reorientació o gir, mentre que les zones grogues/ocres delimiten els patrons de *crawling* o desplaçament lineal.

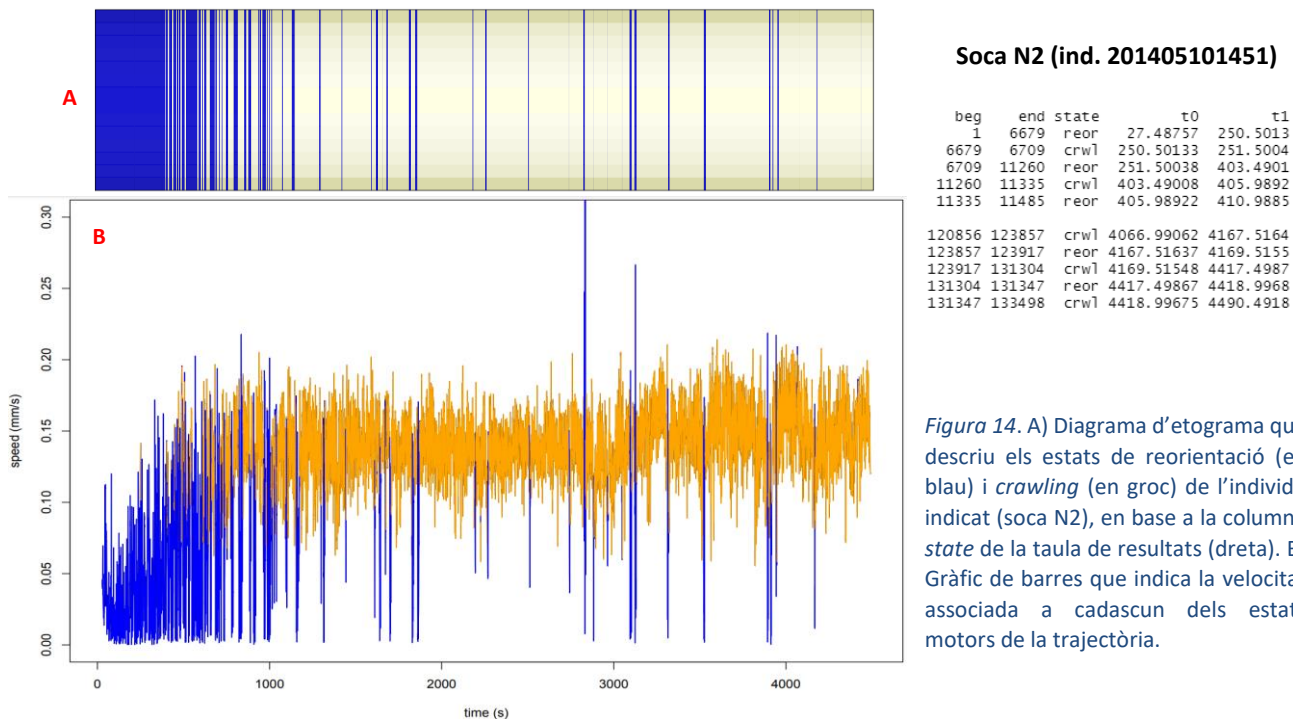


Figura 14. A) Diagrama d'etograma que descriu els estats de reorientació (en blau) i *crawling* (en groc) de l'individu indicat (soca N2), en base a la columna *state* de la taula de resultats (dreta). B) Gràfic de barres que indica la velocitat associada a cadascun dels estats motors de la trajectòria.

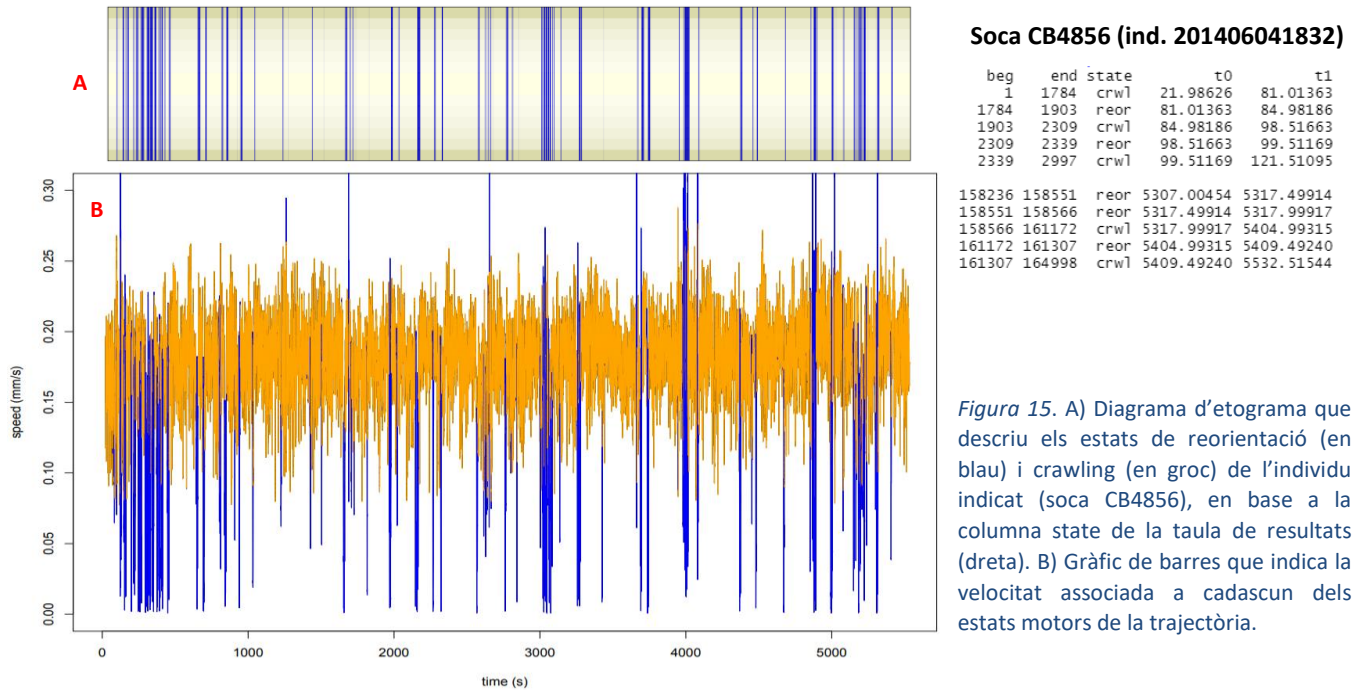


Figura 15. A) Diagrama d'etograma que descriu els estats de reorientació (en blau) i crawling (en groc) de l'individu indicat (soca CB4856), en base a la columna state de la taula de resultats (dreta). B) Gràfic de barres que indica la velocitat associada a cadascun dels estats motors de la trajectòria.

Per últim, les figures 16 i 17 mostren les distribucions de densitat de les velocitats angulars (en radians per segon) per a cadascun dels organismes assajats. En groc es visualitzen els valors associats a un moviment de serpenteig rectilini, i en blau els propis d'un patró de reorientació.

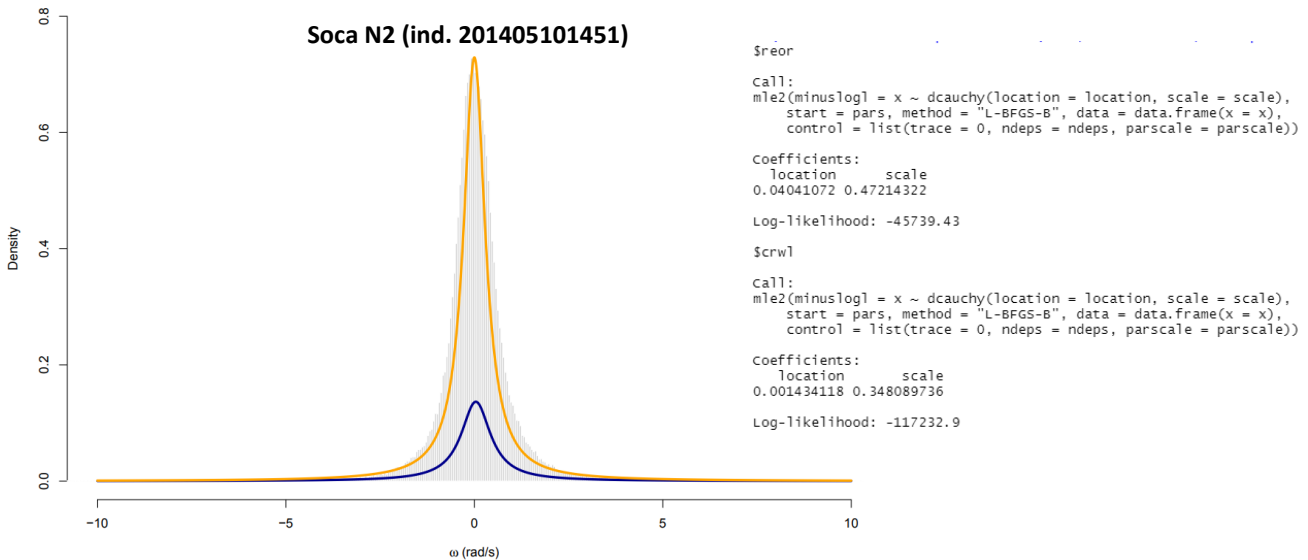


Figura 16. Distribucions de densitat (eix d'ordenades) de les velocitats angulars en rad/s (eix d'abscisses) de l'individu indicat (soca N2), distingint els valors designats al moviment de crawling (ocre) i de reorientació o looping (blau). A la dreta es mostra el model computat de velocitat angular que identifica els dos tipus de patró locomotor.

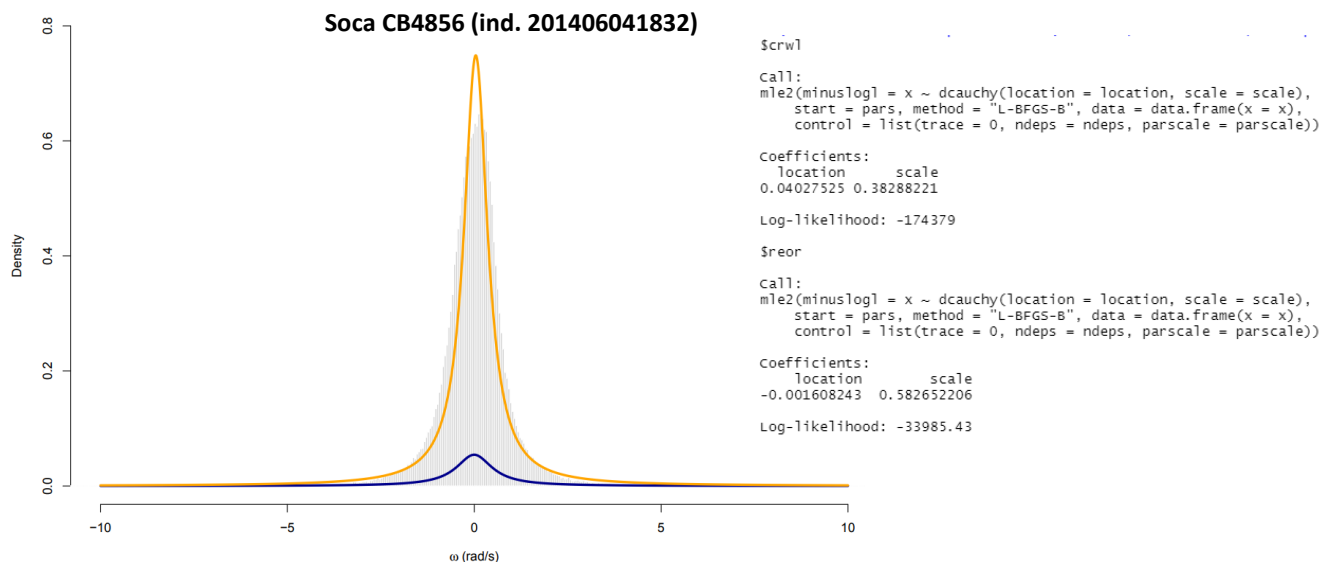


Figura 17. Distribucions de densitat (eix d'ordenades) de les velocitats angulars en rad/s (eix d'abscisses) de l'individu indicat (soca CB456), distingint els valors designats al moviment de crawling (ocre) i de reorientació o looping (blau). A la dreta es mostra el model computat de velocitat angular que identifica els dos tipus de patró locomotor.

A nivell de còmput, s'han estimat els paràmetres de les equacions de Langevin presentades al subapartat E de la secció 4.2.1, tal i com es pot veure a la taula 1. Per a cada individu es detalla el valor de la velocitat mitjana i la seva desviació típica, així com la força de torsió perpetuada al llarg del recorregut i la desviació típica dels angles de gir. També s'especifiquen les unitats de cada paràmetre i el mètode de càlcul per a cadascun.

Taula 1. Resum dels paràmetres de les equacions de Langevin calculats per als individus emprats com a exemple de les dues soques d'estudi.

	Paràmetre	Valor	Unitats	Mètode
Soca N2 (ind. 201405101451)	Velocitat mitjana ( $v_s$ )	0.142008	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02061559	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.0002036562	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.08298545	rad	SD (turning angles)
Soca CB4856 (ind. 201406041832)	Velocitat mitjana ( $v_s$ )	0.1826239	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02425312	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.007355222	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.09744722	rad	SD (turning angles)

## 6 Discussió

### 6.1 Variació del microbioma entèric del gènere *Caenorhabditis*

En el present Treball de Fi de Màster s'ha desenvolupat un meta-anàlisi comparatiu de la microbiota intestinal de diverses espècies del gènere *Caenorhabditis* entre diferents entorns de proliferació de l'hoste. Els resultats presentats a l'apartat anterior posen en evidència la magnitud de l'estudi i el gran ventall d'aproximacions plausibles per a resoldre els diferents objectius plantejats.

Tal i com es pot comprovar amb l'escalat multidimensional no-mètric de la figura 3, les comunitats microbianes semblen agrupar-se homogeniament al llarg del gradient de naturalització considerat per a les condicions de creixement dels nematodes hoste. Les el·lipses traçades s'alineen en concordança, separant les mostres en base al medi de desenvolupament dels cucs processats. De fet, l'escala de categories de proliferació té una transició irrefutable que va de natural a experimental, mantenint la condició de creixement en microcosmos com a satèl·lit a l'alçada de la categoria natural.

A més a més, fa l'efecte que la distribució de les mostres ve influenciada pel substrat de partida on es mantenen els cucs, ja que al voltant de les comunitats intestinals de cada categoria hi apareix un núvol de punts amb les mostres del microbioma del sòl corresponent a la mateixa condició. Això podria explicar per què la proliferació en microcosmos es distingeix del creixement silvestre, ja que el substrat emprat en els estudis d'on procedeixen els *datasets* d'aquella condició (principalment *Berg\_assembly* i *Berg\_host*) foren pre-tractats abans d'iniciar el manteniment i cultiu dels nematodes. Per tant, les característiques del sòl d'origen on els metazous s'han desenvolupat influeixen significativament en la composició dels microbiomes intestinals. Cal tenir present que *Caenorhabditis* és un metazou terrestre que ingereix la matèria que troba reptant pel substrat, per tant tot el que mengi pot tenir un impacte directe sobre la identitat de la seva microbiota. Ara bé, existeix un component de diferenciació considerable entre la microbiota del sòl i la microbiota intestinal d'una mateixa condició, ja que els punts són pròxims però no es solapen completament.

Així mateix, el grau de dispersió de les comunitats intestinals és menor a mesura que la condició de creixement està més controlada. Per exemple, les mostres naturals dels diferents *datasets* s'agrupen bé sota la mateixa categoria, però la varietat d'entorns de mostreig provoca que l'el·lipse que les conté tingui un major radi ja que la riquesa d'espècies serà també molt més alta. En canvi, les mostres de microbioma experimental romanen molt més compactes donat que en l'estudi de procedència es subministra una microbiota concreta, de 12 o 14 tàxons, per colonitzar els nematodes. Es tracta, doncs, d'una situació molt més controlada que condueix a un menor grau de dispersió de les dades.

D'altra banda, l'escalat NMDS de la figura 4 mostra què succeeix quan les condicions de creixement homòlogues es condensen i es suprimeixen les mostres de substrat. De la mateixa manera que a la figura anterior, s'aprecia un gradient de naturalització de les mostres que permet estudiar la dispersió dels microbiomes amb independència dels substrats. De fet, la distribució resulta equivalent a l'observada en la figura 3 (mostres naturals més disperses, microcosmos molt diferenciats i mostres experimentals compactades en una petita regió). En aquest cas, la simplificació de les categories i l'eliminació dels sòls permet contemplar encara millor les diferències existents entre la comunitat intestinal dels nematodes desenvolupats en entorns diferents.

Així mateix, el gradient de naturalització es segueix mantenint en l'escalat NMDS de la figura 5. En aquest cas, s'observa que l'espècie *C. remanei* proliferada en condicions naturals aboca a una microbiota altament variable (és a dir, amb un alt grau de dispersió, fins i tot major a *C. elegans* en la mateixa situació). En general, però, l'entorn de creixement té una influència molt més marcada que l'espècie de nematode en sí; això s'observa amb les condicions de microcosmos, que es separen novament de la resta d'el·lipses però sense fer una

distinció precisa entre les espècies crescudes en aquest entorn. Això concorda amb els resultats del diagrama de caixes múltiple de la figura 6A, en el qual s'infereix que la dispersió associada a les mostres d'entorn natural és molt més elevada (major distància al centroide) que la d'entorn semi-artificial i experimental, siguin quines siguin les espècies d'estudi. També s'observa que les dades de *C. tropicalis* per a condicions de microcosmos estan poc esbiaixades, però això és degut a la mínima representació que té en els *datasets* de partida. Quelcom similar succeeix amb les mostres de *C. briggsae*. Pel que fa als còmputos del test de Tukey (figura 6B), es pot afirmar que existeixen diferències estadísticament significatives (p-valor ajustat menor a 0.001) entre totes les condicions de creixement de *C. elegans*, així com entre les categories de proliferació silvestre i en microcosmos per a *C. remanei*, demostrant novament que l'efecte del substrat sobre la microbiota intestinal és molt superior al que exerceix l'espècie del nematode hoste.

Quant a la descripció taxonòmica de la comunitat entèrica de *C. elegans* al llarg del gradient de naturalització (figura 7), es detecten diferències notables entre les condicions de creixement avaluades. Això permet tenir una bona traçabilitat de la procedència del nematode, doncs la sobre-representació de determinades famílies taxonòmiques pot ajudar a determinar l'entorn de proliferació del cuc. De forma semblant, la comparació del perfil microbià característic de 3 espècies de *Caenorhabditis* aïllades de l'entorn natural (figura 8) contempla diferències significatives determinades per la pròpia espècie dels individus. Val a dir, però, que *C. remanei* i *C. briggsae* tenen una molt menor representació als metagenomes analitzats, per la qual cosa els resultats d'aquestes dues espècies poden presentar certs biaixos.

Pel que fa a l'avaluació del *core microbiome* natiu de *C. elegans* (figura 9), els gèneres més representatius identificats en el present treball corresponen majoritàriament als filums *Gammaproteobacteria* i *Bacteroidota*, la qual cosa era d'esperar donat que les aproximacions presents a la literatura han demostrat la prevalença d'aquests tàxons a l'intestí del nematode en condicions naturals de creixement<sup>[26,48,49]</sup>. El *core microbiome* de *C. remanei* (també procedent d'una proliferació silvestre) està alhora constituït per membres dels mateixos filums, tot i que en aquest cas s'infereix fins i tot una diversitat major. Destaca en ambdós casos la predominança de les famílies *Enterobacteriaceae*, *Sphingomonadaceae* i *Pseudomonadaceae*, els gèneres de les quals són comensals típics del tracte gastrointestinal de nematodes i mamífers<sup>[26, 46]</sup>. Cal remarcar que l'abundància relativa mitjana escollida com a llindar de selecció, establerta a 0.5%, no és tan baixa com podria semblar, tenint en compte que la gran majoria de tàxons determinats presenten una abundància inferior al 0.1%.

## 6.2 Contrast de motilitat entre diverses soques de *C. elegans*

L'algoritme computacional simplificat i optimitzat que s'ha presentat en l'actual treball ha facilitat la caracterització motora de les soques N2 i CB4856 de *C. elegans*.

Tal i com es pot comprovar amb les trajectòries de la figura 11 (i també valorant les presentades a l'Annex II), existeix una variabilitat interindividual pel que fa al recorregut que segueixen els individus d'una mateixa soca. És més, no existeixen trets distintius en els patrons locomotors visualitzats que permetin distingir ambdues soques de *C. elegans* en base a la seva trajectòria.

Les velocitats lineals que prenen els nematodes al llarg del seu recorregut es poden assignar a una funció de distribució de probabilitat mixta constituïda per una part elemental de tipus gamma (amb les velocitats més baixes, en blau) i una altra de tipus gaussiana (amb les velocitats més altes, en groc) (figures 12 i 13). Ambdues soques configuren, de fet, la mateixa distribució. Si aquesta informació es compara amb la il·lustració de les taules d'etograma (figures 14 i 15), aleshores es revela l'associació intrínseca entre els valors de velocitat (figura 14-15B) i els dos estats motors que pot presentar aquest tipus d'organisme (figura 14-15A): les velocitats de la



distribució gamma constitueixen el moviment de reorientació del cuc (com girs, *loops* i piruetes), mentre que les velocitats de la distribució normal descriuen les fases de serpenteig rectilini o *crawling*. Això té sentit, ja que el cuc ha de disminuir la seva velocitat per poder girar, i és aleshores quan aquests valors queden marcadament atribuïts a una distribució de tipus gamma. Ara bé, el perfil d'aquesta distribució mixta varia lleugerament entre individus, doncs cadascun d'ells seguirà un desplaçament independent del que presentin els altres.

No hi ha discrepàncies significatives entre els conjunts d'individus de les diferents soques, de fet és habitual que tots ells iniciïn la seva trajectòria amb forts moviments de reorientació (color blau) a velocitats baixes, donat que s'han de familiaritzar amb l'entorn del sistema de *tracking*. A mesura que coneixen aquest medi, comencen a accelerar (color groc) fins assolir una velocitat mitjana estable, la qual modulen en funció de les subseqüents reorientacions que prenguin per girar i canviar de direcció.

D'altra banda, la velocitat angular que prenen els individus d'ambdues soques pot classificar-se en una distribució de Lorentz-Cauchy (figures 16 i 17), amb la qual es detecta que la velocitat angular que prenen els individus durant els fenòmens de reorientació és menor a l'observada durant els processos de *crawling*. De nou, els perfils dels 10 cucs són altament similars, de manera que tampoc es pot atribuir cap distinció plausible per aquesta aproximació matemàtica.

Finalment, la parametrització de les equacions de Langevin (taula 1) demostra que *C. elegans* combina períodes d'exploració amb instants de mostreig local, transitant al llarg de la placa seguint un règim certament difusiu. Tots els cucs tenen una força propulsora ( $\beta$ ) en la direcció del desplaçament a la qual es mouen, i que alhora intenten equilibrar amb una velocitat constant estable ( $v_s$ ). A més a més, l'angle de la seva fisonomia corporal indueix al cuc a seguir moviments de serpenteig transitoris, i l'error associat ( $\sigma_\phi$ ) intervé activament en la modelització de les seves trajectòries.

## 7 Conclusions

### 7.1 Conclusions

Amb el present projecte ha estat possible estudiar computacionalment la microbiota entèrica i el comportament motor de *C. elegans*. D'una banda, s'ha generat un *pipeline* bioinformàtic mitjançant el qual s'ha comprovat que existeix una clara diferenciació en la composició de la microbiota intestinal entre les espècies del gènere *Caenorhabditis* mantingudes en diverses condicions de creixement. Concretament, l'entorn del nematode sembla influir de forma majoritària en el seu perfil microbià. Considerar un gradient de naturalització ha permès tenir una percepció global de la dispersió d'aquestes comunitats en funció del substrat de procedència dels nematodes hoste. Així mateix, s'ha vist que la comunitat intestinal d'aquests cucs està constituïda principalment per membres de la família *Enterobacteriaceae* i *Gammaproteobacteria*, tal i com s'ha vist en altres estudis publicats a la literatura i validant així l'aproximació emprada en aquest treball.

D'altra banda, s'ha analitzat el perfil de locomoció de dues soques de *C. elegans* per tal de contrastar la motilitat entre individus. L'algoritme adaptat que s'ha presentat en el projecte ha facilitat el tractament de 10 fitxers de trajectòries gravades amb un sistema de *tracking* de 90 minuts de seguiment. Com a resultat, s'ha vist que hi ha una marcada variabilitat motora entre cucs d'una mateixa soca i també entre soques. Aquests organismes acostumen a desplaçar-se en base a dos tipus de moviment, *crawlings* o reorientacions, les velocitats dels quals s'atribueixen a una distribució elemental particular (normal o gamma, respectivament) que en conjunt defineixen una distribució de densitat de probabilitat mixta. Alhora, ha estat possible parametritzar els principals termes de les equacions del model de Langevin adaptat a la dinàmica animal, els valors dels quals resulten útils per descriure matemàticament la locomoció d'un organisme reptant com *C. elegans*.

En conclusió, la possibilitat de caracteritzar aquests dos aspectes rellevants en la biologia del nematode suposen la base de futurs estudis en els quals s'infereixi una possible interconnexió entre la microbiota intestinal i el moviment de *C. elegans*.

## 7.2 Línies de futur

En primer lloc, l'estudi de la microbiota intestinal de *Caenorhabditis* s'ha dut a terme a partir de dades de diversos *datasets*, els quals poden determinar els resultats obtinguts en aquest projecte donades les seves particularitats intrínseques a nivell experimental. Seria convenient estudiar més a fons el número de metagenomes d'amplicó a incloure en el meta-anàlisi, així com valorar adequadament la profunditat de mostreig de cada *dataset*. És important evitar biaixos a les dades que puguin alterar les conclusions biològiques que s'extrauen d'aquestes.

Per un altre costat, també caldria realitzar diferents aproximacions per tal de caracteritzar la locomoció de diferents espècies del gènere *Caenorhabditis*, ja que no s'han observat diferències entre soques. També s'haurien de computar els paràmetres macroscòpics que permeten validar el model, amb l'objectiu d'estudiar els patrons de difusió presents a la seva trajectòria. De la mateixa manera, tal i com s'ha comentat, convindria determinar quins canvis succeeixen en el comportament animal d'aquest nematode en base a diferències en la seva microbiota entèrica. Això podria ser objecte d'estudi en un futur projecte de tesi.

## 7.3 Seguiment de la planificació

En general, totes les tasques previstes inicialment al pla de treball s'han desenvolupat amb èxit, tot i que ha hagut endarreriments temporals en la consecució d'algunes etapes donada la seva elevada dificultat. No ha calgut introduir plans de contingència i s'han assolit tots els objectius proposats inicialment.

## 8 Glossari

<b>GBA</b>	<i>Gut-Brain Axis</i>
<b>rDNA</b>	<i>ribosomal Desoxyribonucleic Acid</i>
<b>NCBI</b>	<i>National Center for Biotechnology Information</i>
<b>GNU</b>	<i>GNU's not Unix</i>
<b>OTU</b>	<i>Operational Taxonomic Unit</i>
<b>IDE</b>	<i>Integrated Development Environment</i>
<b>NCBI – SRA</b>	<i>Sequence Read Archive</i>
<b>MG-RAST</b>	<i>Metagenomic Rapid Annotations using Subsystems Technology</i>
<b>EBI – ENA</b>	<i>European Bioinformatics Institute – European Nucleotide Archive</i>
<b>CEAB-CSIC</b>	Centre d'Estudis Avançants de Blanes - Consell Superior d'Investigacions científiques
<b>BLAST</b>	<i>Basic Local Alignment Search Tool</i>
<b>CPU</b>	<i>Central Processing Unit</i>
<b>RAM</b>	<i>Random access memory</i>
<b>EE</b>	<i>Expected Error</i>
<b>NGS</b>	<i>Next-Generation Sequencing</i>
<b>RNA</b>	<i>Ribonucleic Acid</i>
<b>SINA</b>	<i>SILVA Incremental Aligner</i>
<b>LCA</b>	<i>Last Common Ancestor</i>
<b>CSV</b>	<i>Comma-Separated Values</i>
<b>NMDS</b>	<i>Non-Metric MultiDimensional Scaling</i>
<b>NGM</b>	<i>Nematode Growth Medium</i>
<b>LB</b>	<i>Lysogeny Broth</i>

## Bibliografia

1. Nicholas, W. (1975). *The biology of free-living nematodes*. Oxford University Press. Oxford, England.
2. Palikaras, K., Tavernarakis, N. (2013). *Brenner's Encyclopedia of Genetics: Caenorhabditis elegans (Nematode)*, 2<sup>nd</sup> Edition, vol. 1, 404–408. doi:10.1016/b978-0-12-374984-0.00186-8
3. Kaletta, T., Hengartner, M. O. (2006). Finding function in novel targets: *C. elegans* as a model organism. *Nature reviews - Drug discovery*, 5(5), 387–398. <https://doi.org/10.1038/nrd2031>
4. Nigon, V. M., Félix, M. A. (2017). History of research on *C. elegans* and other free-living nematodes as model organisms. *WormBook: The Online Review of C. Elegans Biology*, 1-84. <https://doi.org/10.1895/wormbook.1.181.1>
5. Girard, L. R., Fiedler, T. J., Harris, T. W., et al. (2007). WormBook: the online review of *Caenorhabditis elegans* biology. *Nucleic acids research*, 35 (Database issue), D472–D475. <https://doi.org/10.1093/nar/gkl894>
6. Corsi A.K., Wightman B., Chalfie M. (2015). A transparent window into biology: A primer on *Caenorhabditis elegans*. *WormBook: The Online Review of C. Elegans Biology*, doi/10.1895/wormbook.1.177.1. URL: [http://www.wormbook.org/chapters/www\\_celegansintro/celegansintro.html](http://www.wormbook.org/chapters/www_celegansintro/celegansintro.html)
7. Meneely, P. M., Dahlberg, C. L., Rose, J. K. (2019). Working with worms: *Caenorhabditis elegans* as a model organism. *Current Protocols Essential Laboratory Techniques*, 19, e35. doi: 10.1002/cpet.35
8. Douglas, A. E. (2019). Simple animal models for microbiome research. *Nature Reviews Microbiology*, 17(12), 764–775. <https://doi.org/10.1038/s41579-019-0242-1>
9. Zhang, S., Li, F., Zhou, T., et al. (2020). *Caenorhabditis elegans* as a Useful Model for Studying Aging Mutations. *Front. Endocrinol.*, 11:554994. <https://doi.org/10.3389/fendo.2020.554994>
10. Attene-Ramos, M.S. (2014). *Encyclopedia of Toxicology: High Throughput Screening*. (), 916–917. doi:10.1016/b978-0-12-386454-3.00209-8
11. Bain, O., Baldwin, J. G., Beveridge, I., et al. (2013). *Handbook of Zoology: Gastrotricha, Cycloneuralia and Gnathifera*. Vol.2 *Nematoda*. A. Schmidt-Rhaesa ed. De Gruyter Publisher.
12. Brusca, R. C. (2016). *Invertebrates* (3rd ed.). Oxford University Press.
13. Oswald, M. (2020). *Caenorhabditis elegans*. Animal Diversity Web. URL: [https://animaldiversity.org/accounts/Caenorhabditis\\_elegans/](https://animaldiversity.org/accounts/Caenorhabditis_elegans/)
14. Riddle DL., Blumenthal T., Meyer BJ., et al. (1997) Introduction to *C. elegans*: *C. elegans II*. 2nd edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. Chapter 1. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK20183/>
15. Morais, L. H., Schreiber, H. L., Mazmanian, S. K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology*, 19(4), 241–255. <https://doi.org/10.1038/s41579-020-00460-0>
16. Cryan, J. F., O’Riordan, K. J., Cowan, C. S. M., et al. (2019). The Microbiota-Gut-Brain Axis. *Physiological Reviews*, 99(4), 1877–2013. <https://doi.org/10.1152/physrev.00018.2018>
17. Rieder, R., Wisniewski, P. J., Alderman, B. L., et al. (2017). Microbes and mental health: A review. *Brain, Behavior, and Immunity*, 66, 9–17. <https://doi.org/10.1016/j.bbi.2017.01.016>
18. Adak, A., & Khan, M. R. (2019). An insight into gut microbiota and its functionalities. *Cellular and Molecular Life Sciences*, 76(3), 473–493. <https://doi.org/10.1007/s00018-018-2943-4>
19. Liu, S., Gao, J., Zhu, M., et al. (2020). Gut Microbiota and Dysbiosis in Alzheimer’s Disease: Implications for Pathogenesis and Treatment. *Molecular Neurobiology*, 57(12), 5026–5043. <https://doi.org/10.1007/s12035-020-02073-3>
20. Sarkar, A., Harty, S., Lehto, S. M., et al. (2018). The Microbiome in Psychology and Cognitive Neuroscience. *Trends in Cognitive Sciences*, 22(7), 611–636. <https://doi.org/10.1016/j.tics.2018.04.006>
21. Roubalová, R., Procházková, P., Papežová, H., et al. (2020). Anorexia nervosa: Gut microbiota-immunebrain interactions. *Clinical Nutrition*, 39(3), 676–684. <https://doi.org/10.1016/j.clnu.2019.03.023>
22. Gomaa, E. Z. (2020). Human gut microbiota/microbiome in health and diseases: A review. *Antonie van Leeuwenhoek*, 113(12), 2019–2040. <https://doi.org/10.1007/s10482-020-01474-7>
23. Cuevas-Sierra, A., Ramos-Lopez, O., Riezu-Boj, J. I., et al. (2019). Diet, Gut Microbiota, and Obesity: Links with Host Genetics and Epigenetics and Potential Applications. *Advances in Nutrition*, 10(Suppl 1), S17–S30. <https://doi.org/10.1093/advances/nmy078>
24. Long-Smith, C., O’Riordan, K. J., Clarke, G., et al. (2020). Microbiota-Gut-Brain Axis: New Therapeutic Opportunities. *Annual Review of Pharmacology and Toxicology*, 60(1), 477–502. <https://doi.org/10.1146/annurev-pharmtox-010919-023628>

25. Rinninella, E., Raoul, P., Cintoni, M., *et al.* (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms*, 7(1). <https://doi.org/10.3390/microorganisms7010014>
26. Zhang F., Berg M., Dierking K., *et al.* (2017). *Caenorhabditis elegans* as a Model for Microbiome Research. *Front. Microbiol.* 8:485. doi: 10.3389/fmicb.2017.00485
27. Wang HX., Wang YP. (2016). Gut Microbiota-brain Axis. *Chin Med J*, 129, 2373-2380. <https://doi.org/10.4103/0366-6999.190667>
28. Berg, M., Stenuit, B., Ho, J., *et al.* (2016). Assembly of the *Caenorhabditis elegans* gut microbiota from diverse soil microbial environments. *The ISME Journal*, 10(8), 1998-2009. doi:10.1038/ismej.2015.253
29. Bartumeus, F., Campos, D., Ryu, W. S., *et al.* (2016). Foraging success under uncertainty: search tradeoffs and optimal space use. *Ecology letters*, 19(11), 1299–1313. <https://doi.org/10.1111/ele.12660>
30. Salvador, L. C., Bartumeus, F., Levin, S. A., & Ryu, W. S. (2014). Mechanistic analysis of the search behaviour of *Caenorhabditis elegans*. *Journal of the Royal Society, Interface*, 11(92), 20131092. <https://doi.org/10.1098/rsif.2013.1092>
31. Bell WJ. (1991). *Searching behaviour: the behavioural ecology of finding resources*. Springer Netherlands - Chapman and Hall Animal Behavioural Series. London, UK.
32. Bartumeus F., da Luz MGE., Viswanathan GM., Catalan J. (2005). Animal search strategies: a quantitative random-walk analysis. *Ecology*, 86, 3078 – 3087. doi:10.1890/04-1806
33. Méndez, V., Campos, D., Bartumeus, F. (2014). *Stochastic Foundations in Movement Ecology: Random search strategies*. Springer Series in Synergetics (ed. Haken, Herman.). Springer Berlin Heidelberg, Berlin, Germany, pp. 177–205.
34. Bartumeus F. (2007). Levy processes in animal movement: an evolutionary hypothesis. *Fractals*, 15, 151– 162. doi:10.1142/S0218348X07003460
35. Bartumeus F., Levin SA. (2008). Fractal reorientation clocks: linking animal behaviour to statistical patterns of search. *Proc. Natl Acad. Sci.*, 105(49), pp. 19072 – 19077. doi:10.1073/pnas.0801926105
36. Nathan, R., Getz, WM., Revilla, E., *et al.* (2008). A movement ecology paradigm for unifying organismal movement research. *Proc. Natl Acad. Sci.*, 105(49), pp. 19052 – 19059. doi:10.1073/pnas.0800375105
37. Stephens, G. J., Johnson-Kerner, B., Bialek, W., Ryu, W. S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS computational biology*, 4(4), e1000028. <https://doi.org/10.1371/journal.pcbi.1000028>
38. Marizzoni, M., Gurry, T., Provasi, S., *et al.* (2020). Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Faecal Samples. *Frontiers in Microbiology*, 11(), 1262. doi:10.3389/fmicb.2020.01262
39. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460-2461. <https://doi.org/10.1093/bioinformatics/btq461>
40. Rognes, T., Flouri, T., Nichols, B., *et al.* (2016). VSEARCH: a versatile open-source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
41. Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature methods*, 10(10), 996–998. <https://doi.org/10.1038/nmeth.2604>
42. R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: <https://www.R-project.org/>.
43. RStudio Team. (2019). RStudio: Integrated Development for R. *RStudio, Inc.*, Boston, MA URL: <http://www.rstudio.com/>
44. Michelot, T., Gloaguen, P., Blackwell, PG., *et al.* (2019). The Langevin diffusion as a continuous-time model of animal movement and habitat selection. *Methods in Ecology and Evolution*, 2041-210X.13275–. doi:10.1111/2041-210x.13275
45. Gerbaba, T. K., Green-Harrison, L., & Buret, A. G. (2017). Modeling Host-Microbiome Interactions in *Caenorhabditis elegans*. *Journal of nematology*, 49(4), 348–356. <https://doi.org/10.21307/jofnem-2017-082>
46. Backes, C., Martínez-Martínez, D., Cabreiro, F. (2021). *C. elegans*: A biosensor for host–microbe interactions. *Lab Anim* 50, 127–135. <https://doi.org/10.1038/s41684-021-00724-z>
47. Berg M., Zhou XY., Shapira M. (2016) Host-Specific Functional Significance of *Caenorhabditis* Gut Commensals. *Front. Microbiol.* 7:1622. doi: 10.3389/fmicb.2016.01622
48. Johnke, J., Dirksen, P., Schulenburg, H. (2020). Community assembly of the native *C. elegans* microbiome is influenced by time, substrate and individual bacterial taxa. *Environmental Microbiology*, 22(4), 1265–1279. doi:10.1111/1462-2920.14932.
49. Dirksen, P., Marsh, SA., Braker, I. *et al.* (2016). The native microbiome of the nematode *Caenorhabditis elegans*: gateway to a new host-microbiome model. *BMC Biology*, 14:24. doi: 10.1186/s12915-016-0258-1.

50. Samuel, B., Rowedder, H., Braendle, C. *et al.* (2016). *Caenorhabditis elegans* responses to bacteria from its natural habitats. *PNAS*, vol. 113 nº 27. <https://doi.org/10.1073/pnas.1607183113>
51. Tsibidis, G.D., Tavernarakis, N. (2007). *Nemo*: a computational tool for analyzing nematode locomotion. *BMC Neurosci* 8, 86. <https://doi.org/10.1186/1471-2202-8-86>
52. Koopman, M., Peter, Q., Seinstra, R.I., *et al.* (2020). Assessing motor-related phenotypes of *Caenorhabditis elegans* with the wide field-of-view nematode tracking platform. *Nat Protoc* 15, 2071–2106. <https://doi.org/10.1038/s41596-020-0321-9>
53. Moy K., Li W., Tran H.P., *et al.* (2015) Computational Methods for Tracking, Quantitative Assessment, and Visualization of *C. elegans* Locomotory Behavior. *PLoS ONE* 10(12): e0145870. <https://doi.org/10.1371/journal.pone.0145870>
54. Husson, S. J. *et al.* (2012). Keeping track of worm trackers, *WormBook: The Online Review of C. Elegans Biology*. doi/10.1895/wormbook.1.156.1. [http://www.wormbook.org/chapters/www\\_tracking/tracking.html](http://www.wormbook.org/chapters/www_tracking/tracking.html)
55. Cronin, C. J., Mendel, J. E., Mukhtar, S., *et al.* (2005). An automated system for measuring parameters of nematode sinusoidal movement. *BMC genetics*, 6, 5. <https://doi.org/10.1186/1471-2156-6-5>
56. Likitlersuang, J., Stephens, G., Palanski, K., *et al.* (2012). *C. elegans* Tracking and Behavioral Measurement. *J. Vis. Exp.* (69), e4094, doi: [10.3791/4094](https://doi.org/10.3791/4094).
57. Schretter, C. E., Vielmetter, J., Bartos, I., *et al.* (2018). A gut microbial factor modulates locomotor behaviour in *Drosophila*. *Nature*, 563(7731), 402–406. <https://doi.org/10.1038/s41586-018-0634-9>
58. Shu, R., Hahn, D. A., Jurkevitch, E., *et al.* (2021). Sex-Dependent Effects of the Microbiome on Foraging and Locomotion in *Drosophila suzukii*. *Frontiers in Microbiology*, 12. <https://www.frontiersin.org/article/10.3389/fmicb.2021.656406>
59. Ericsson, A.C., Hart, M.L., Kwan, J. *et al.* Supplier-origin mouse microbiomes significantly influence locomotor and anxiety-related behavior, body morphology, and metabolism. *Commun Biol* 4, 716 (2021). <https://doi.org/10.1038/s42003-021-02249-0>
60. Tatham, S., Lanes, A., Harris, B. *et al.* (1999). *PuTTY: a free SSH and Telnet client - PuTTY Licence*. [online]. URL: <https://www.chiark.greenend.org.uk/~sgtatham/putty/>
61. Dirksen, P., Assié, A., Zimmermann, J. *et al.* (2020). CeMbio - The *Caenorhabditis elegans* Microbiome Resource. *G3: Genes | Genomes | Genetics*, vol. 20, 3025-3039. doi: <https://doi.org/10.1534/g3.120.401309>
62. Taylor, L. J., Abbas, A., Bushman, F. D. (2020). *grabseqs*: Simple downloading of reads and metadata from multiple next-generation sequencing data repositories. *Bioinformatics*, 36(11), p. 3607-3609. <https://doi.org/10.1093/bioinformatics/btaa167>
63. SourceForge Team, w/ Tim Kosse. (2001). *FileZilla: The free FTP solution*. [online]. URL: [https://filezilla-project.org/filezilla\\_pro.php](https://filezilla-project.org/filezilla_pro.php)
64. Usearch v11 (Drive5). (n/d). *Merging (assembly) of paired reads*. Data de consulta: 25 de març de 2022. En línia. URL: [https://www.drive5.com/usearch/manual/merge\\_pair.html](https://www.drive5.com/usearch/manual/merge_pair.html)
65. Usearch v11 (Drive5). (n/d). *Ultra-fast search*. Data de consulta: 25 de març de 2022. En línia. URL: [https://drive5.com/usearch/features\\_search.html](https://drive5.com/usearch/features_search.html)
66. Usearch v11 (Drive5). (n/d). *fastq\_mergpairs options*. Data de consulta: 27 de març de 2022. En línia. URL: [https://www.drive5.com/usearch/manual/merge\\_options.html](https://www.drive5.com/usearch/manual/merge_options.html)
67. Usearch v11 (Drive5). (n/d). *fastx\_get\_sample\_names command*. Data de consulta: 27 de març de 2022. En línia. URL: [https://drive5.com/usearch/manual/cmd\\_fastx\\_get\\_sample\\_names.html](https://drive5.com/usearch/manual/cmd_fastx_get_sample_names.html)
68. Rognes, T; *et al.* (2019). *User Commands - VSEARCH User Guide*. [online]. URL: <https://github.com/torognes/vsearch>
69. Illumina, Inc. (2011). Quality Scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred quality scoring. *Technical Note: Sequencing - Illumina*. [online]. URL: [https://www.illumina.com/documents/products/technotes/technote\\_Q-Scores.pdf](https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf)
70. Edgar, R.C., Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* (Oxford, England), 31(21), 3476–3482. <https://doi.org/10.1093/bioinformatics/btv401>
71. Edgar, RC. (2011). *USEARCH Software and documentation – Version 5.2*. [online]. URL: <http://drive5.com/usearch/UsearchUserGuide5.2.pdf>
72. Usearch v5 (Drive5). (n/d). *substring dereplication using version 5.2*. Data de consulta: 29 de març de 2022. [online]. URL: [https://drive5.com/usearch/manual8.1/derep\\_substr\\_v5.html](https://drive5.com/usearch/manual8.1/derep_substr_v5.html)
73. Usearch v11 (Drive5). (n/d). *sortbysize command*. Data de consulta: 29 de març de 2022. [online]. URL: [https://www.drive5.com/usearch/manual/cmd\\_sortbysize.html](https://www.drive5.com/usearch/manual/cmd_sortbysize.html)

74. Sokal, R.R., Sneath, P. (1973). *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. 1st Edition, W. H. Freeman, San Francisco.
75. Usearch v11 (Drive5). (n/d). *Defining and interpreting OTUs*. Data de consulta: 29 de març de 2022. [online]. URL: <https://drive5.com/usearch/manual/otus.html>
76. Ashelford, K., Chuzhanova, N., Fry, J., et al. (2005). At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. *Applied and Environmental Microbiology*, 71(12). doi: <https://doi.org/10.1128/AEM.71.12.7724-7736.2005>
77. Huse, S. M., Welch, D. M., Morrison, H. G., et al. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental microbiology*, 12(7), 1889–1898. <https://doi.org/10.1111/j.1462-2920.2010.02193.x>
78. Usearch v11 (Drive5). *Generating OTUs and ZOTUs*. (2010). Data de consulta: 1 d'abril de 2022. En línia. URL: [https://drive5.com/usearch/manual/pipe\\_otus.html](https://drive5.com/usearch/manual/pipe_otus.html)
79. Usearch v11 (Drive5). (n/d). *UPARSE-OTU algorithm*. Data de consulta: 2 d'abril de 2022. [online]. URL: [https://drive5.com/usearch/manual/uparseotu\\_algo.html](https://drive5.com/usearch/manual/uparseotu_algo.html)
80. Usearch v11 (Drive5). (n/d). *cluster\_otus command*. Data de consulta: 2 d'abril de 2022. [online]. URL: [https://drive5.com/usearch/manual/cmd\\_cluster\\_otus.html](https://drive5.com/usearch/manual/cmd_cluster_otus.html)
81. Usearch v11 (Drive5). (n/d). *OTU table*. Data de consulta: 4 d'abril de 2022. [online]. URL: [https://drive5.com/usearch/manual/otu\\_table.html](https://drive5.com/usearch/manual/otu_table.html)
82. Usearch v11 (Drive5). (n/d). *Creating an OTU table*. Data de consulta: 4 d'abril de 2022. [online]. URL: [https://drive5.com/usearch/manual/pipe\\_otutab.html](https://drive5.com/usearch/manual/pipe_otutab.html)
83. Quast C., Pruesse E., Yilmaz P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.*, 41(D1): D590-D596. <https://doi.org/10.1093/nar/gks1219>
84. The SILVA development Team. (n/d). *SSU and LSU Ref NR datasets (non-redundant)*. Data de consulta: 4 d'abril de 2022. [online]. URL: <https://www.arb-silva.de/projects/ssu-ref-nr/>
85. Pruesse, E., Peplies, J., Glöckner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28(14), 1823-1829. <https://doi.org/10.1093/bioinformatics/bts252>
  - Pruesse, E. (2012). *SINA: high throughput multiple sequence alignment – Command Line Usage v.1.2.9*. [online]. URL: [https://www.arb-silva.de/no\\_cache/download/archive/SINA/1.2.11/](https://www.arb-silva.de/no_cache/download/archive/SINA/1.2.11/)
86. Oksanen, J; Simpson, G; Blanchet, F; et al. (2022). *vegan: Community Ecology Package*. R package version 2.6-2. <https://CRAN.R-project.org/package=vegan>
87. Stiernagle, T. (2006). Maintenance of *C. elegans*. *WormBook: The Online Review of C. Elegans Biology*, doi/10.1895/wormbook.1.101.1. [http://www.wormbook.org/chapters/www\\_strainmaintain/strainmaintain.html](http://www.wormbook.org/chapters/www_strainmaintain/strainmaintain.html)
88. Hart, A. C. (2006). Behavior. *WormBook: The Online Review of C. Elegans Biology*, doi/10.1895/wormbook.1.87.1. URL: [http://www.wormbook.org/chapters/www\\_behavior/behavior.html](http://www.wormbook.org/chapters/www_behavior/behavior.html)
89. Lewis, J. A., Fleming, J. T. (1995). *Caenorhabditis elegans: Modern Biological Analysis of an Organism* | Chapter 1: Basic culture methods. *Methods in cell biology*, 48, 3–29. doi: 10.1016/S0091-679X(08)61381-3
90. Porta-de-la-Riva, M., Fontrodona, L., Villanueva, A., et al. (2012). Basic *Caenorhabditis elegans* methods: synchronization and observation. *Journal of visualized experiments (JoVE)*, 64, e4019. <https://doi.org/10.3791/4019>
91. The Eclipse Foundation. (2022). *IDE and Tools – Desktop IDEs*. [online]. URL: <https://www.eclipse.org/ide/>
92. MacNulty, D. R., Mech, L. D., Smith, D. W. (2007). A Proposed Ethogram of Large-Carnivore Predatory Behavior, Exemplified by the Wolf. *Journal of Mammalogy*. 88 (3): 595–605. doi:10.1644/06-MAMM-A-119R1.1
93. Serway, R. A., Jewett, J. W. (2004). *Physics for Scientists and Engineers* (6ª edició). Thompson-Brooks/Cole. Belmont, CA. <https://archive.org/details/physicssciengv2p00serw/page/n5/mode/2up>
94. Riley, K. F., Hobson, M. P., Bence, S. J. (2006). *Mathematical Methods for Physics and Engineering* (3 ed.). Cambridge University Press. Cambridge, UK. p.1333. [https://archive.org/details/mathematicalmethod00rile\\_192](https://archive.org/details/mathematicalmethod00rile_192)
95. Racca E., Porporato A. (1999). Langevin equations from time series. *Phys Rev E.*, 71(2): 027101–027103. doi: 10.1103/PhysRevE.71.027101.
96. Kawasaki, K. (1973). Simple derivations of generalized linear and nonlinear Langevin equations. *Journal of Physics A: Mathematical, Nuclear and General*, 6(9), 1289–1295. doi: 10.1088/0305-4470/6/9/004.
97. Lloret, R., Campos, D., Ryu, W., et al. (2021). *DYNS3BIO - International Conference on Dynamics in Systems and Synthetic Biology (Abstracts Book): Search behaviour in a model organism: a walk on the wild side of diffusion*. Centre de Recerca Matemàtica (CRM). [online]. URL: [https://www.crm.cat/wp-content/uploads/2021/06/DYNS3BIO\\_IT\\_Booklet.pdf](https://www.crm.cat/wp-content/uploads/2021/06/DYNS3BIO_IT_Booklet.pdf)

98. Bartumeus, F., Catalan, J., Viswanathan, G.M., *et al.* (2008). The influence of turning angles on the success of non-oriented animal searches. *Journal of Theoretical Biology*, 252(1), 43–55. doi: 10.1016/j.jtbi.2008.01.009.

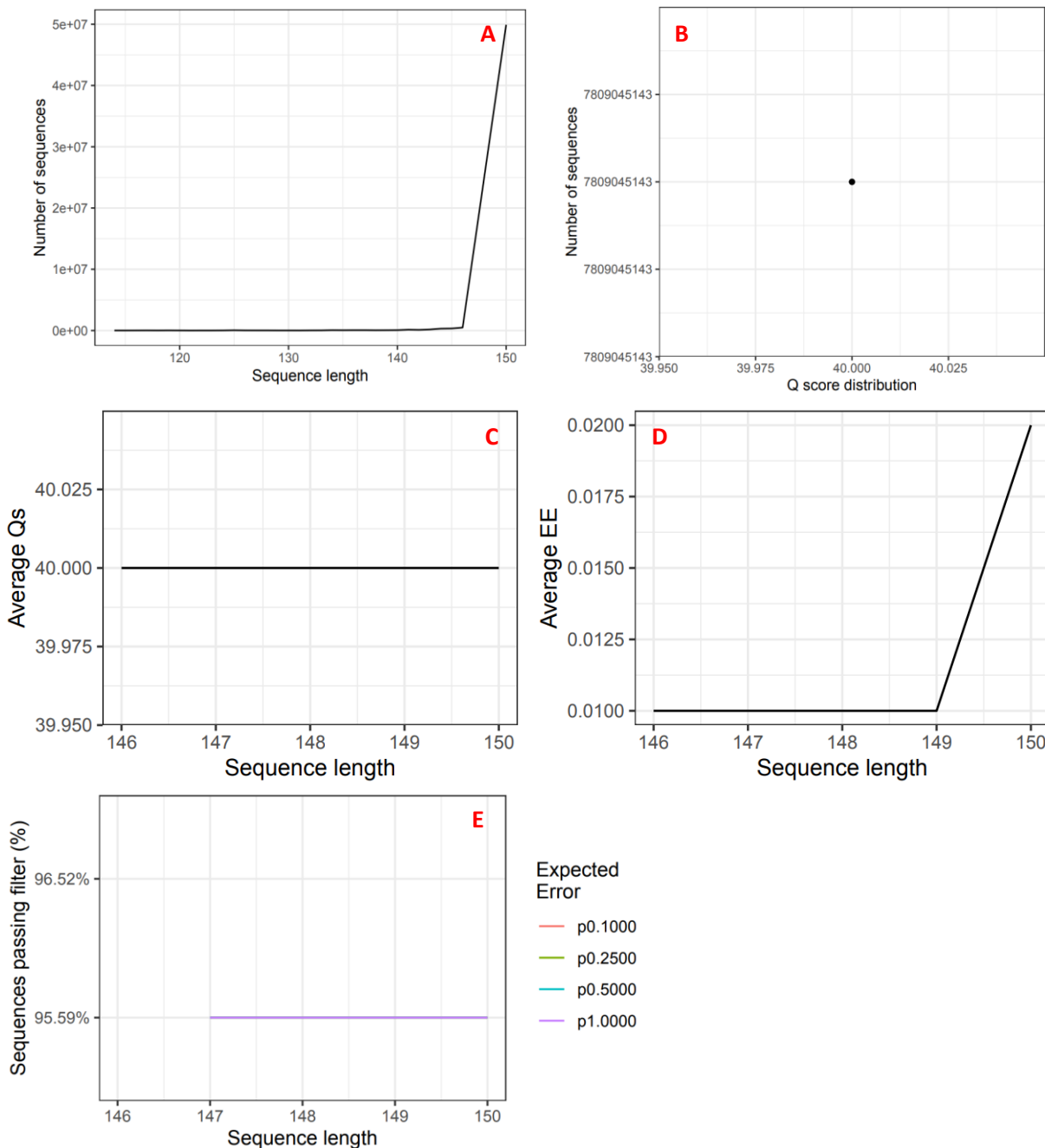


# Annexes

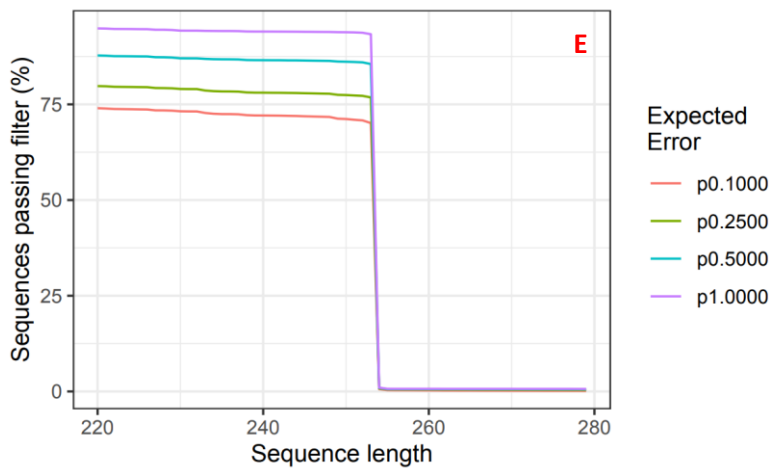
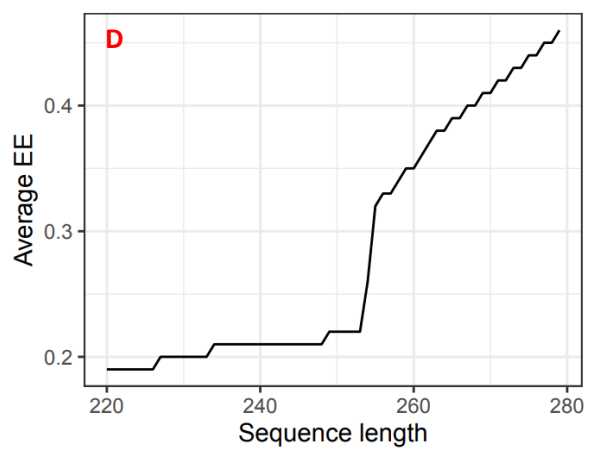
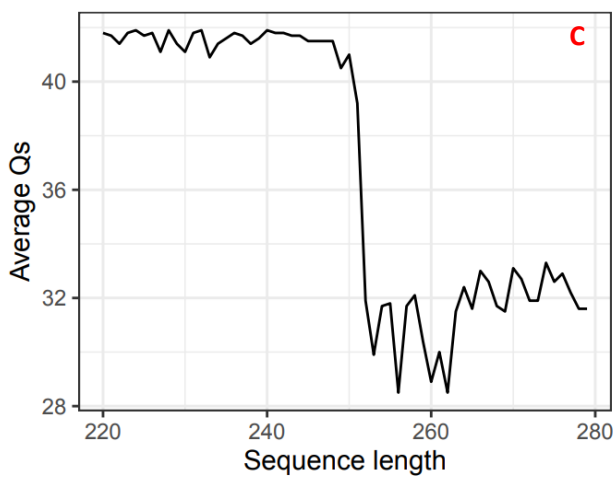
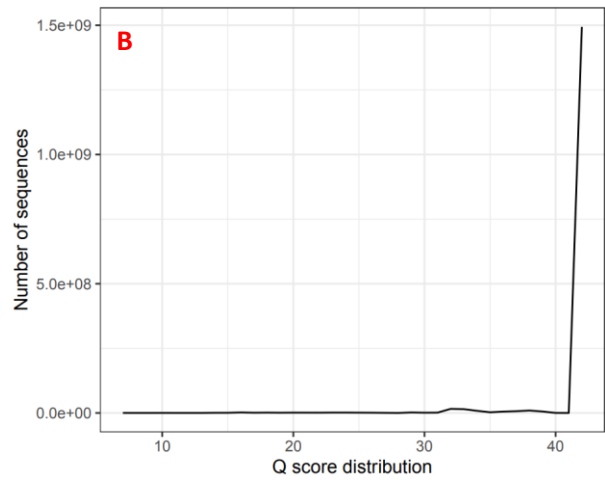
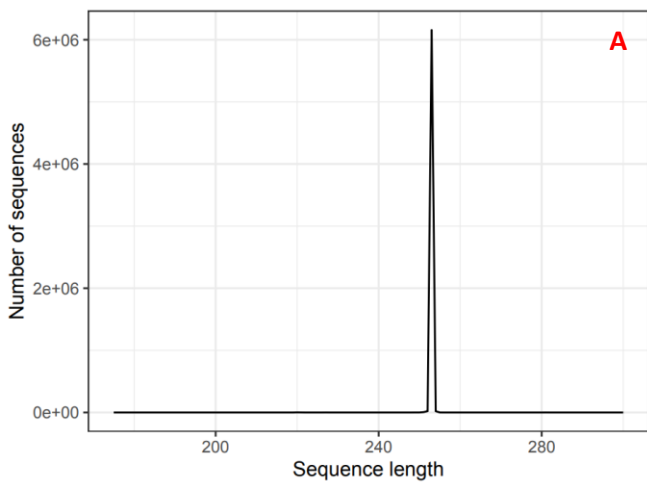
## ANNEX I: Gràfics de qualitat dels *datasets*

En aquest annex es recullen els gràfics de qualitat computats en R a partir dels fitxers d'estadístiques obtinguts per a cadascun dels metagenomes d'amplicó avaluats al meta-anàlisi, tal i com s'ha exemplificat per al cas concret de *Dirksen\_native* en el subapartat C de la secció 4.1.1. Els conjunts de dades presentats s'indiquen segons el nom simplificat amb el qual s'identifiquen en el present treball.

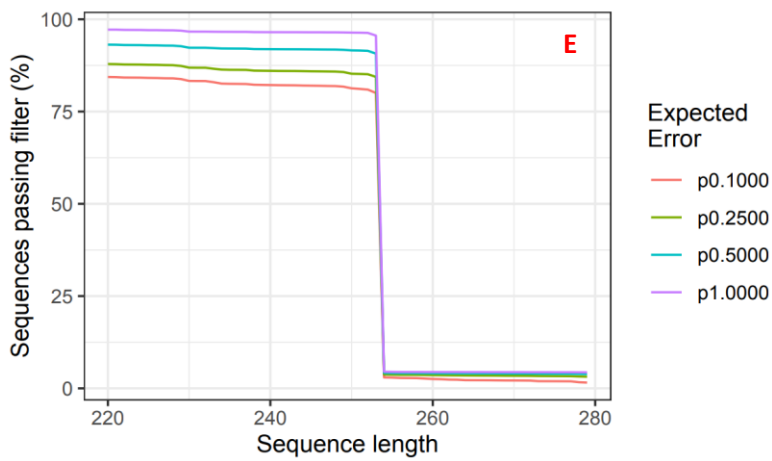
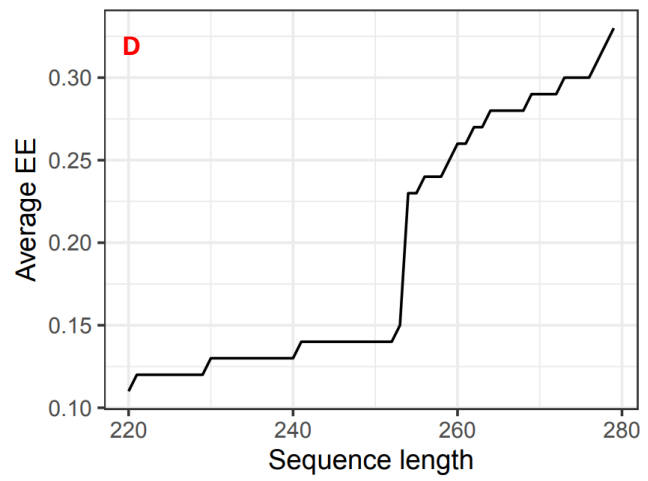
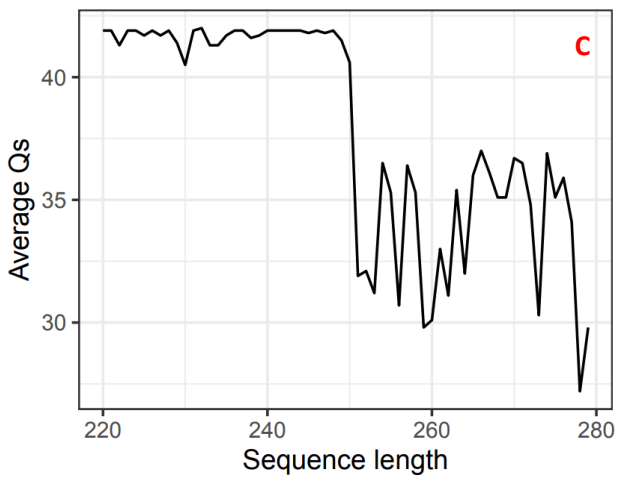
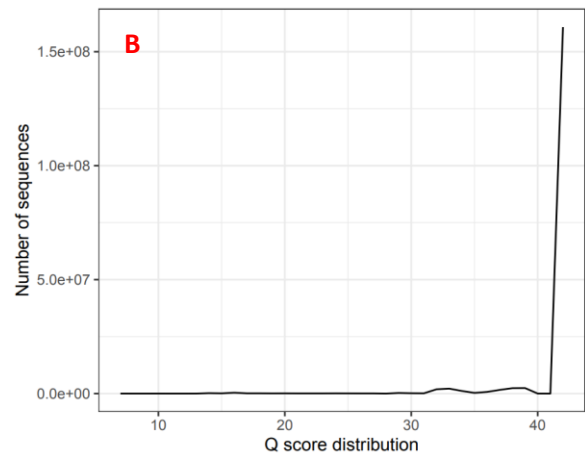
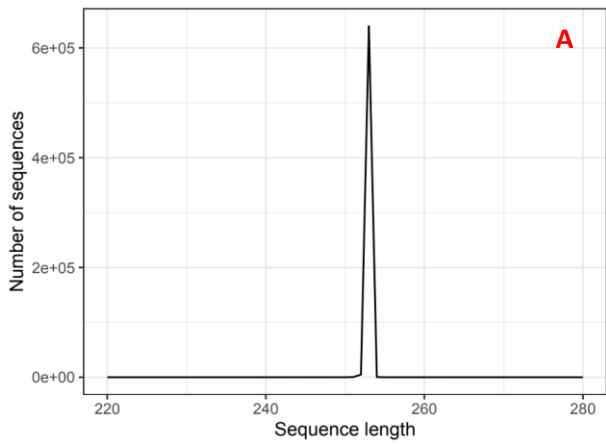
### Berg Assembly



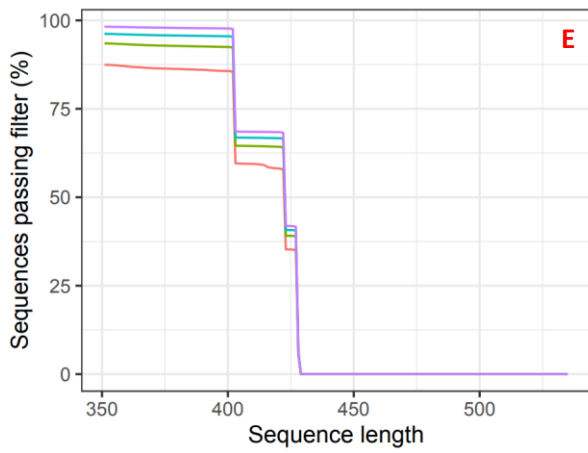
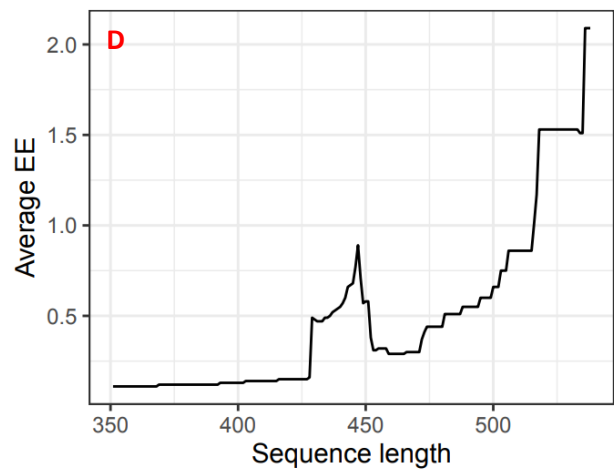
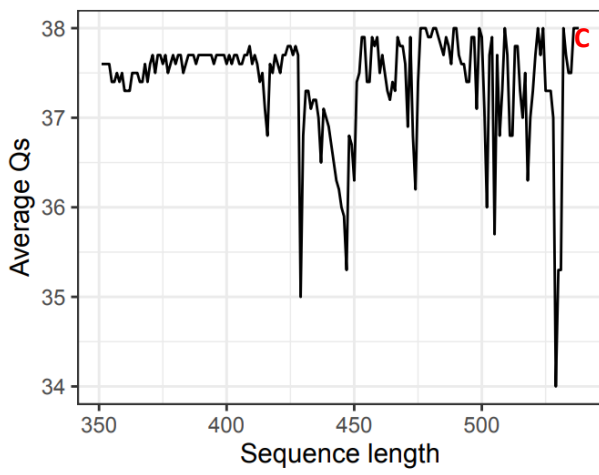
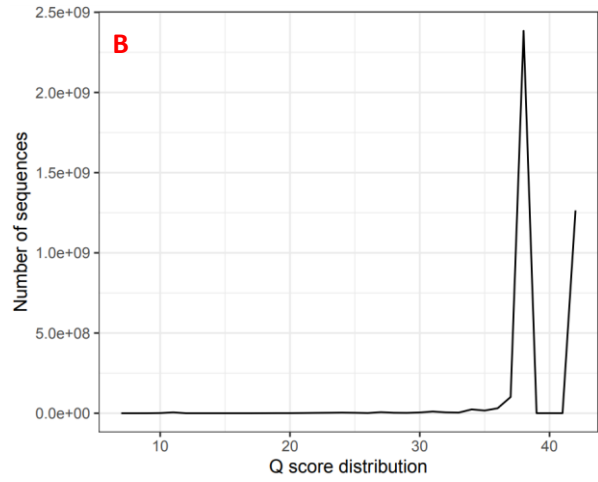
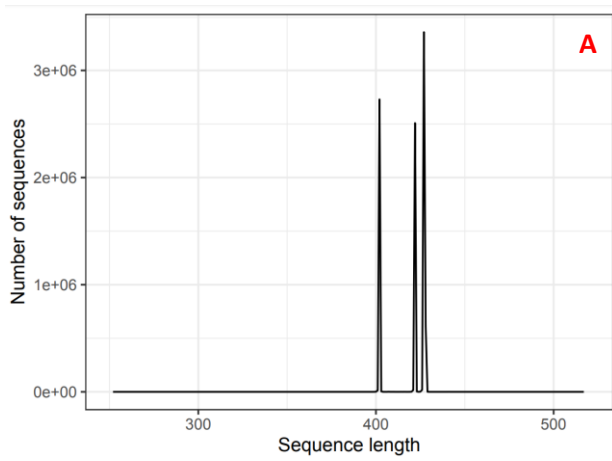
**Dirksen experimental**



**Dirksen CeMbio\_exp2**



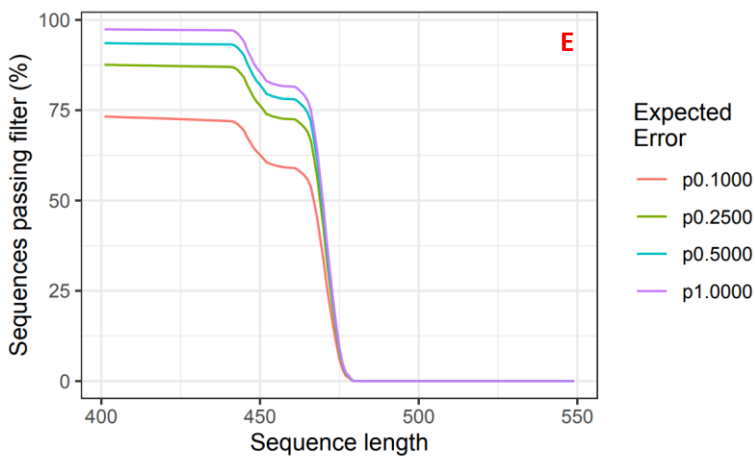
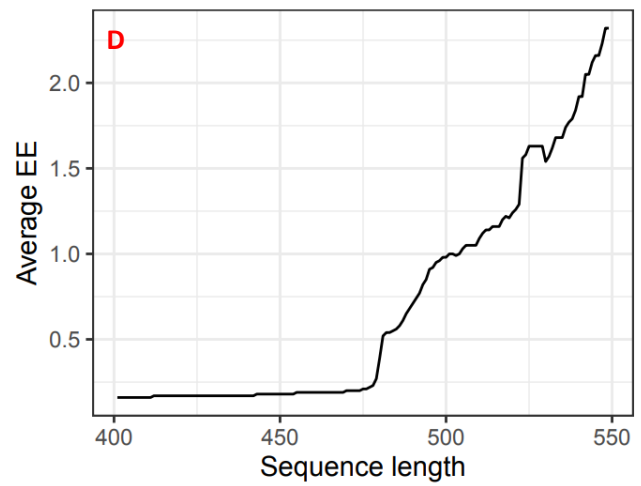
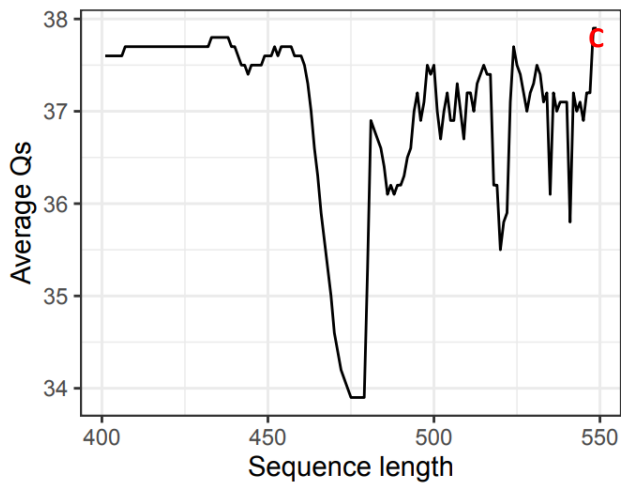
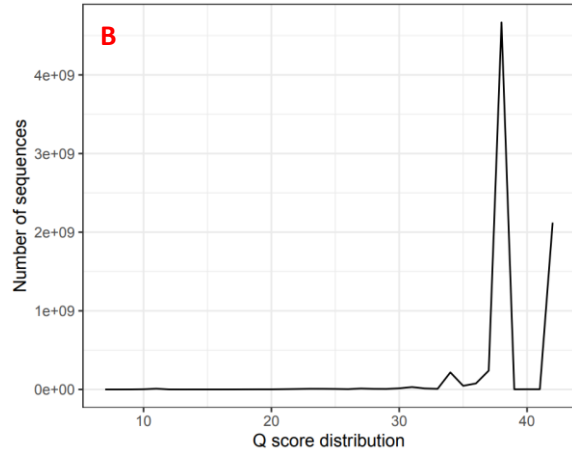
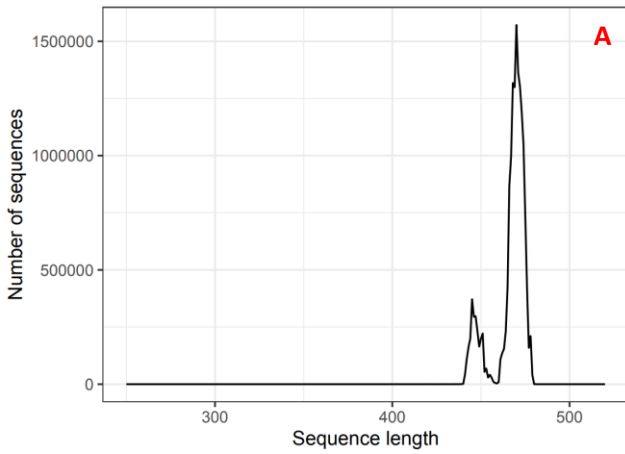
***Dirksen CeMbio\_exp3***



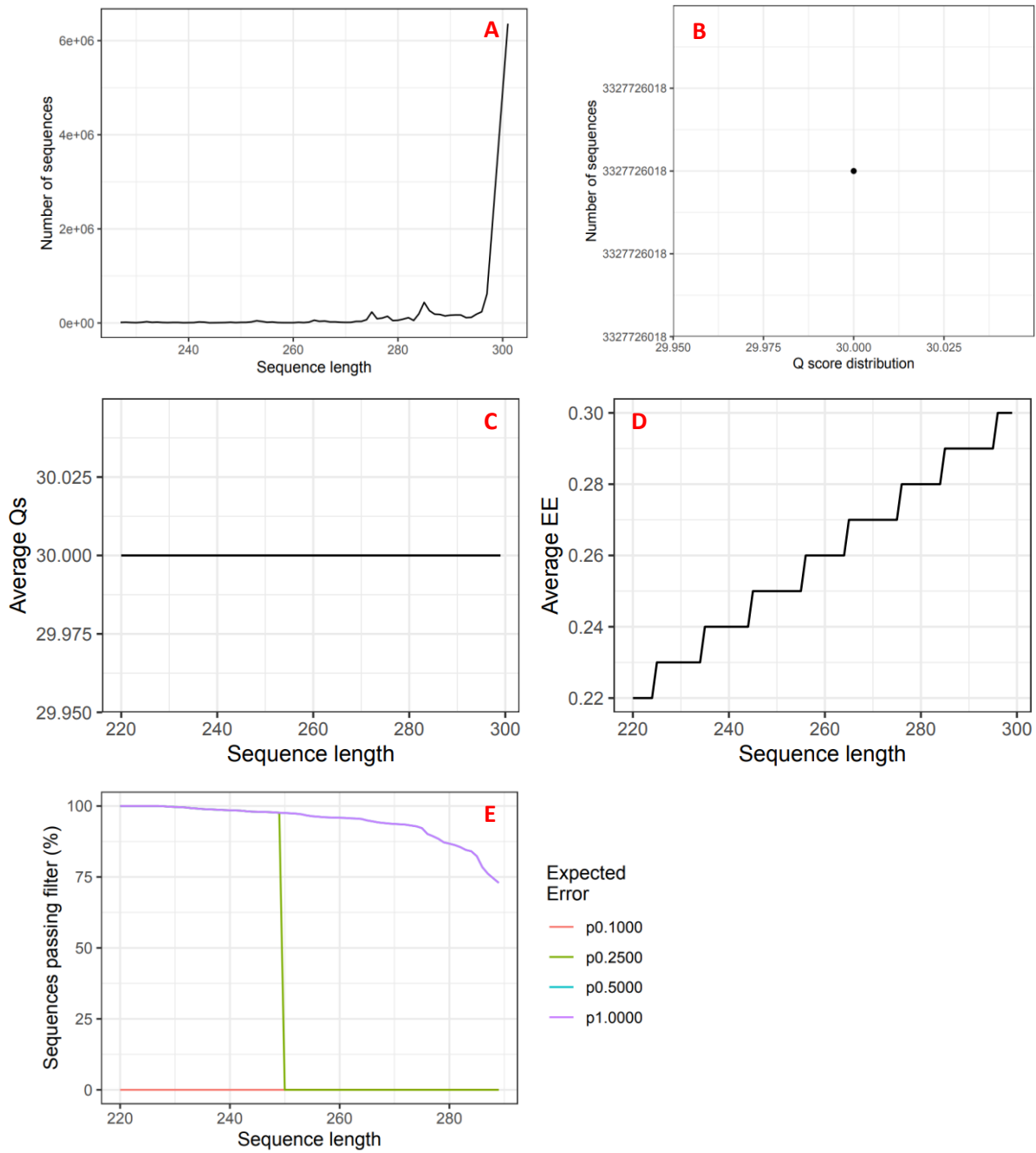
Expected Error

- p0.1000
- p0.2500
- p0.5000
- p1.0000

**Johnke**



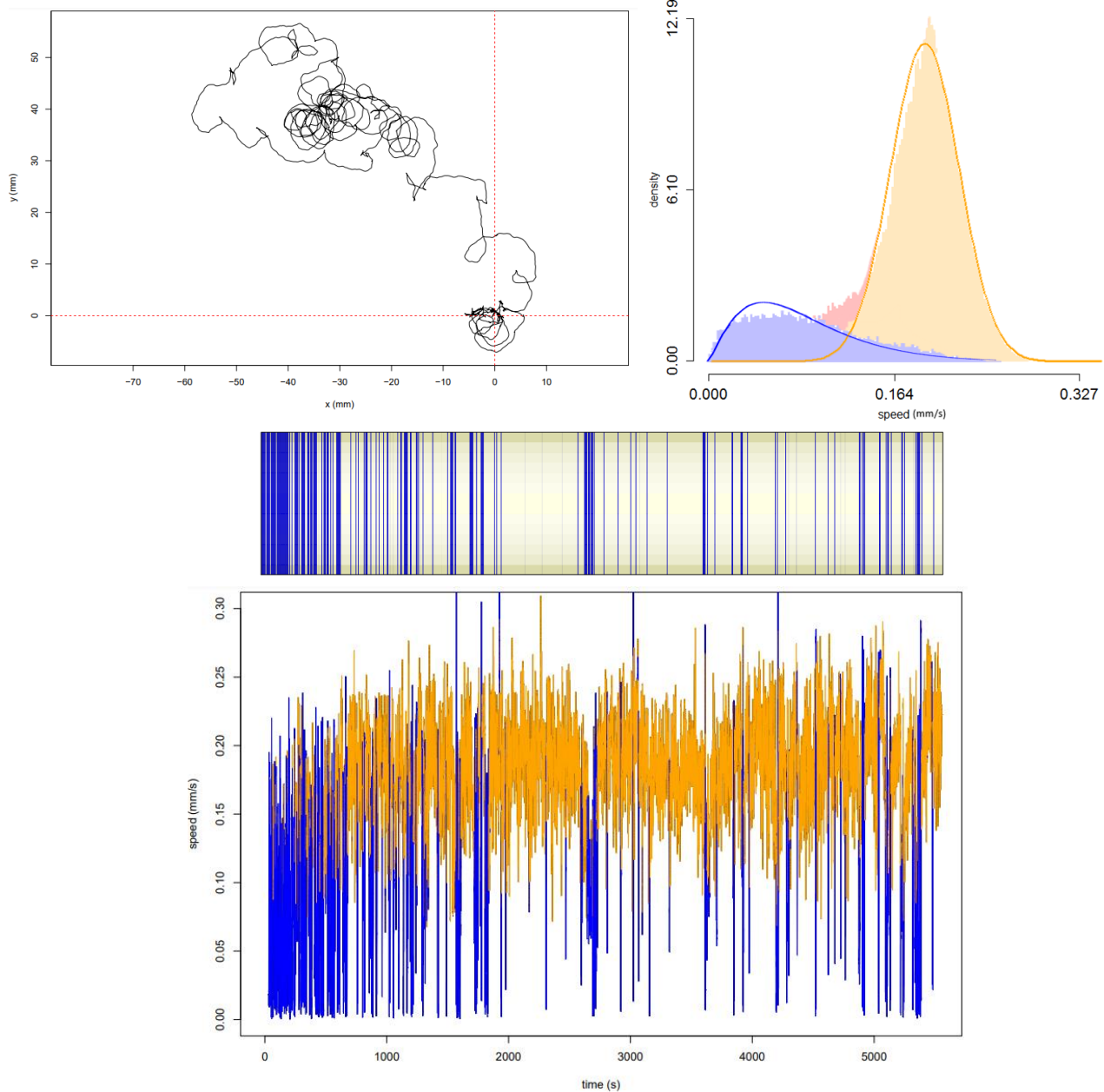
**Samuel**

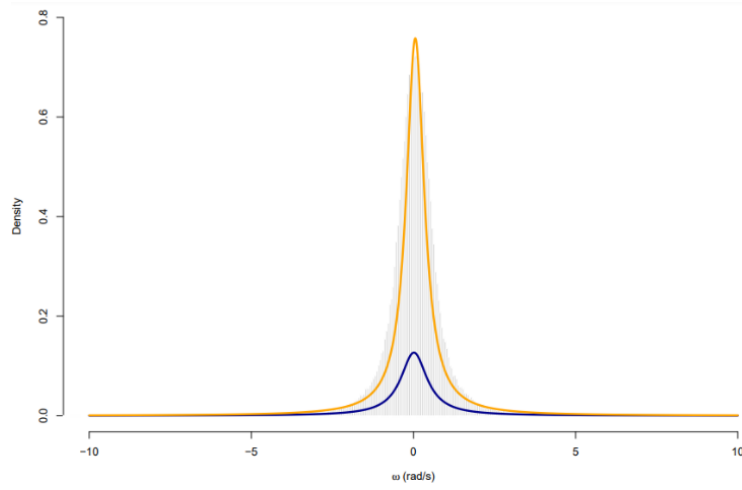


## ANNEX II: Gràfics de locomoció de la resta d'individus

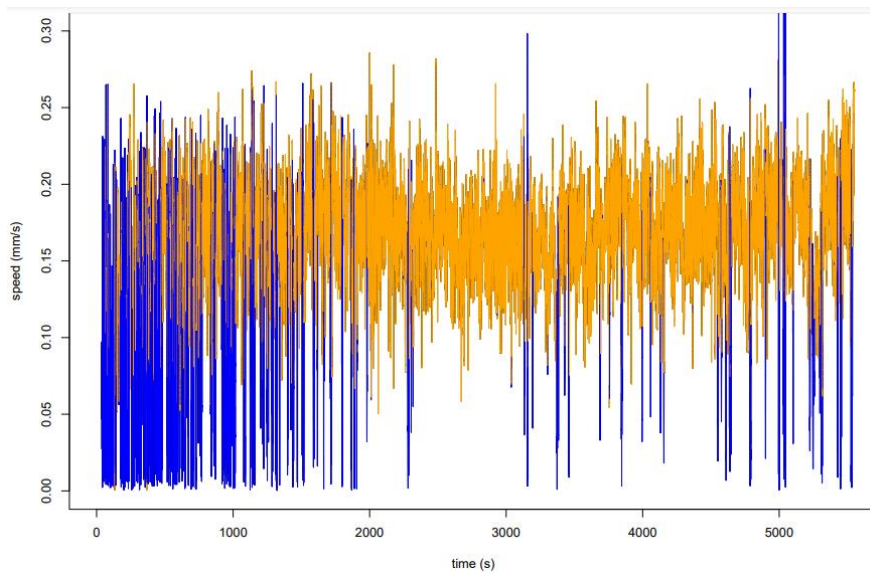
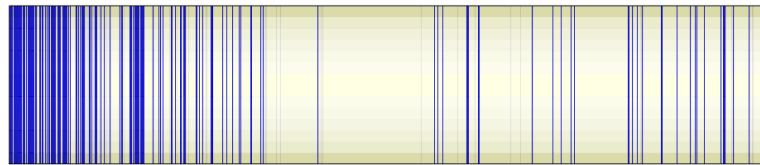
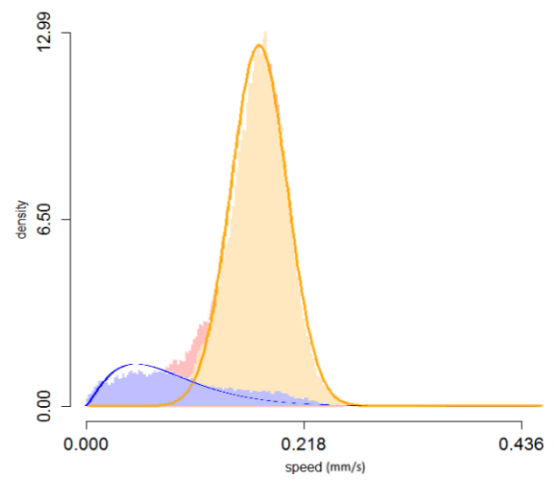
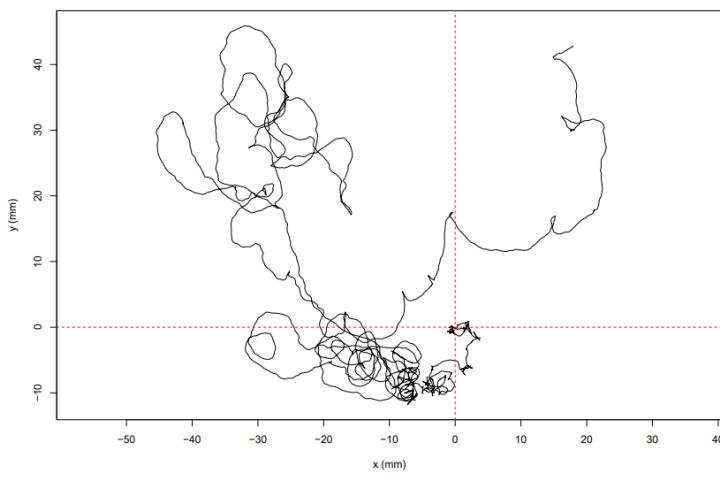
El segon annex inclòs en el present treball està constituït per tot el conjunt de resultats que no han estat inclosos a la secció 5.2, corresponent a la caracterització dels patrons motors de 10 individus de *C. elegans* (5 de la soca N2 i 5 de la soca CB4856). A continuació es detallen resumidament els gràfics obtinguts per als 8 individus restants, indicant la soca corresponent i l'individu al que pertanyen.

### Soca N2 (ind. 201405091556)

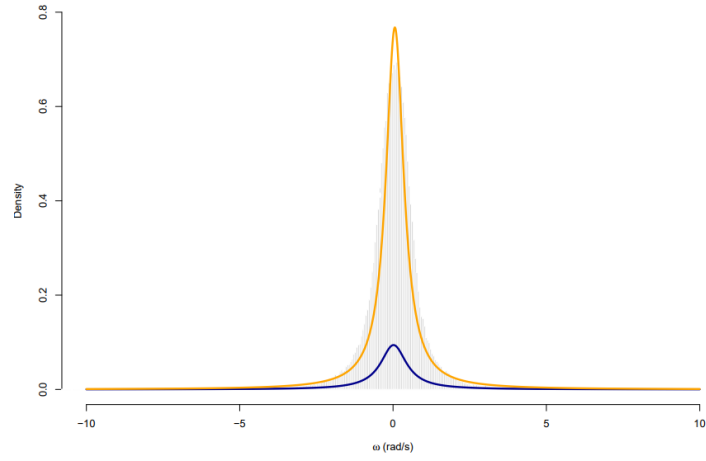




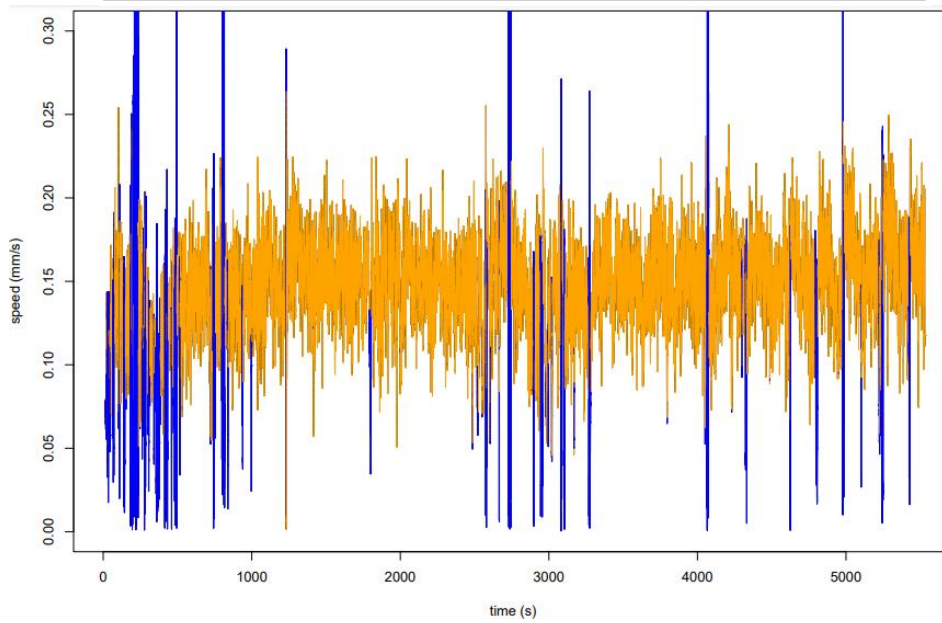
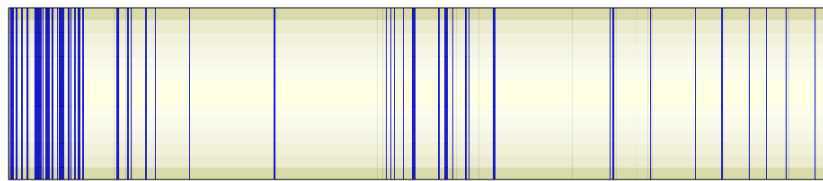
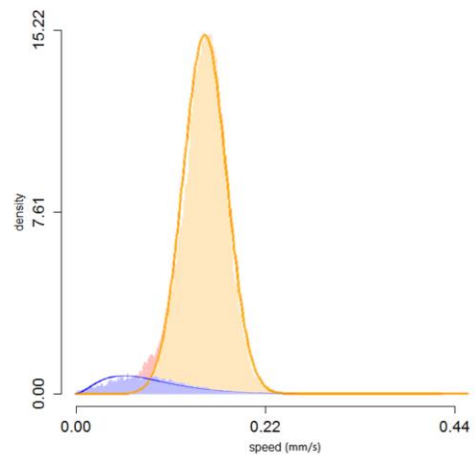
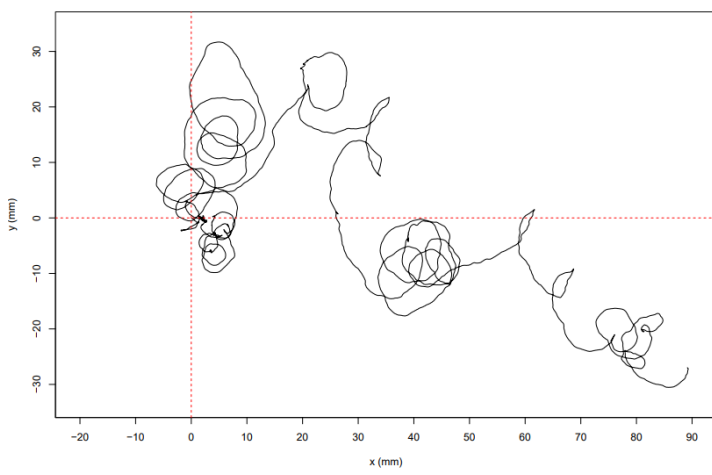
**Soca N2 (ind. 201405091756)**

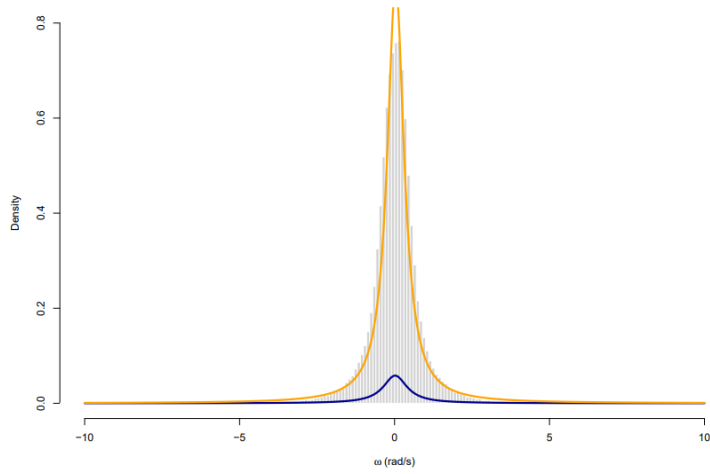




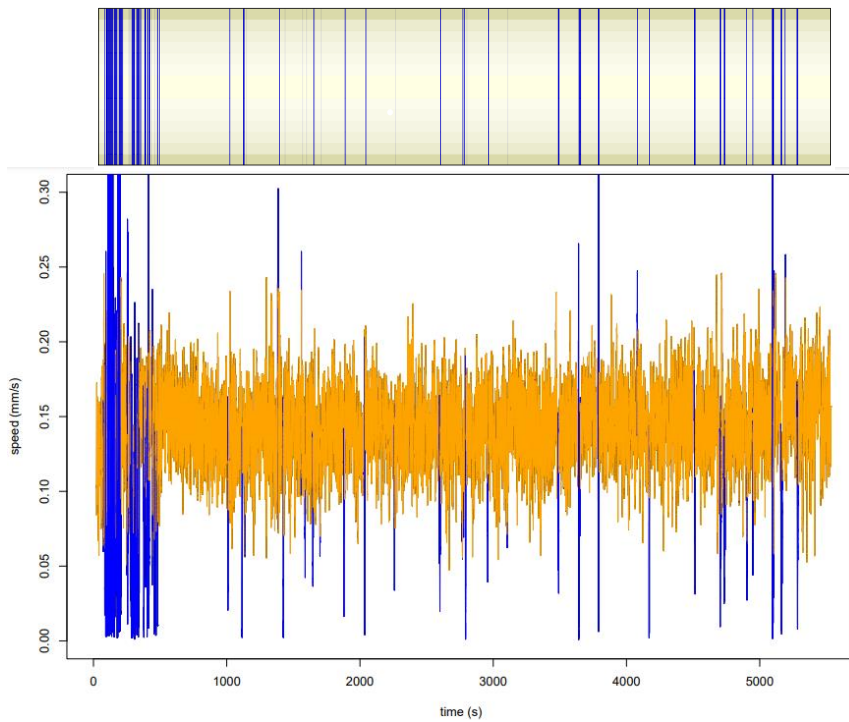
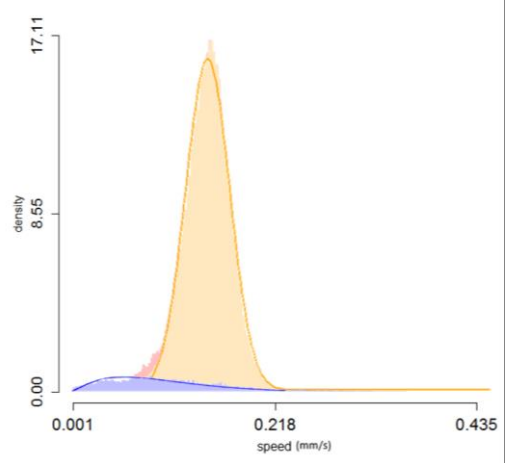
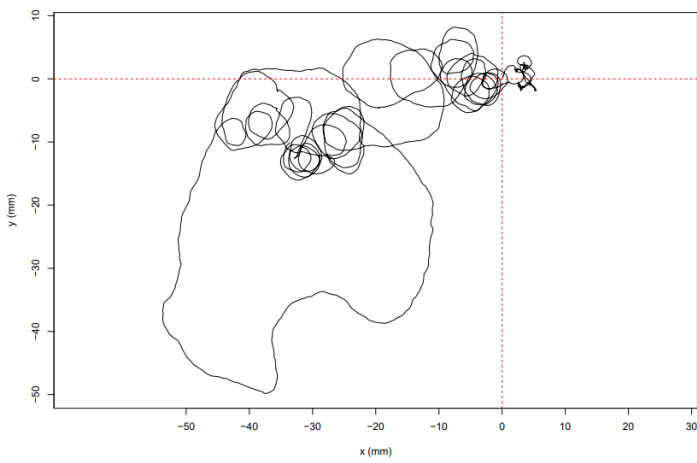


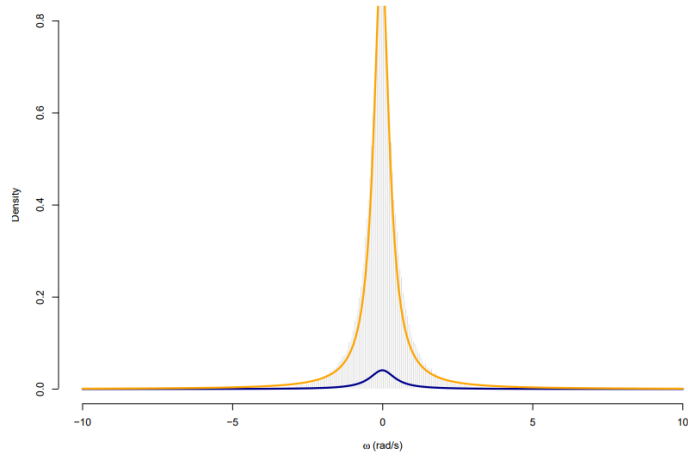
**Soca N2 (ind. 201405101638)**



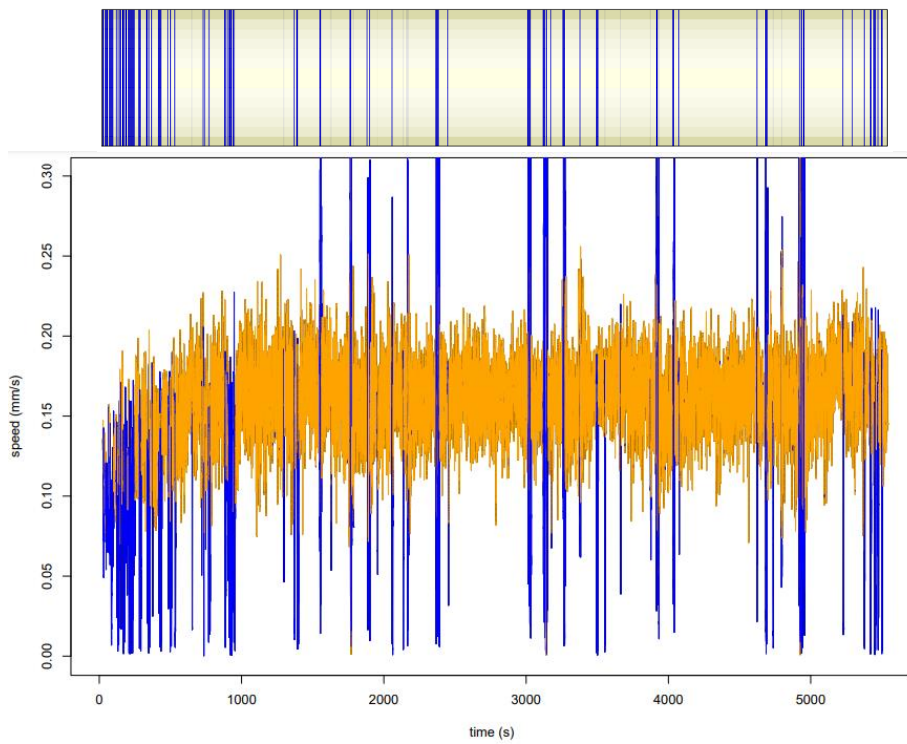
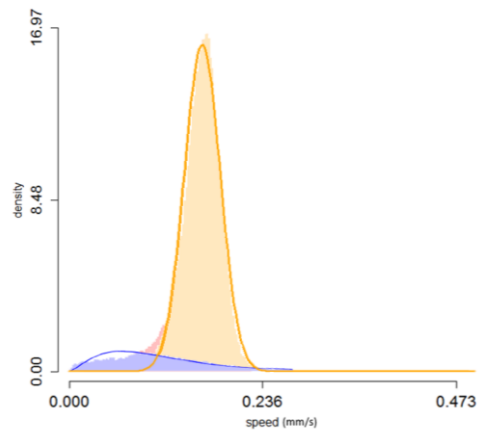
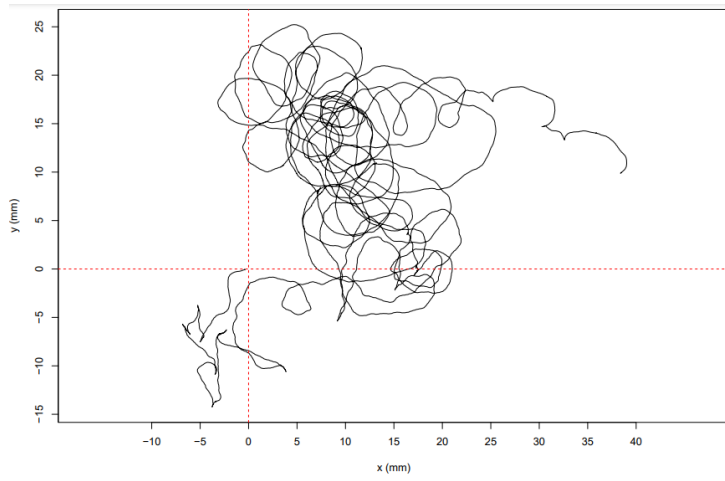


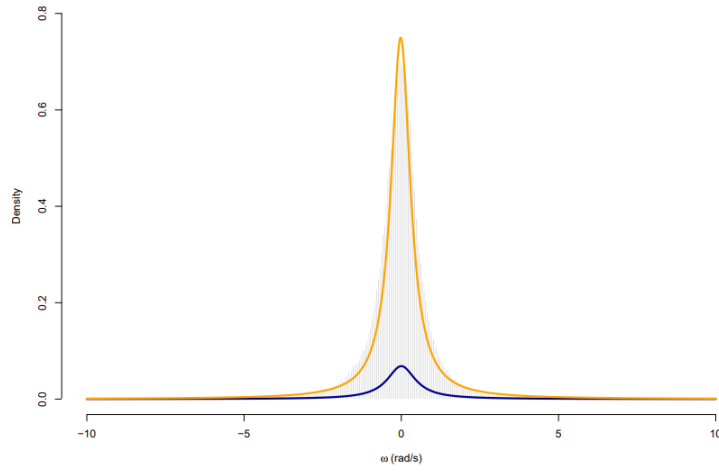
**Soca N2 (ind. 201405120011)**



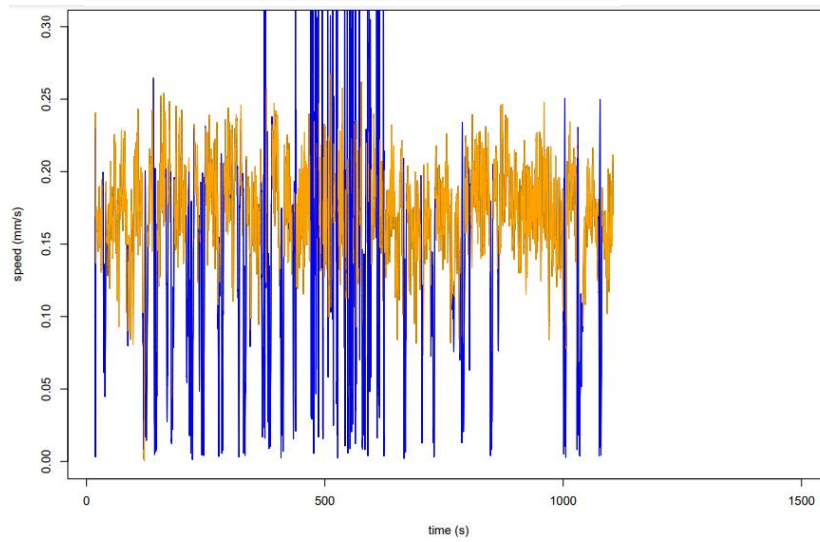
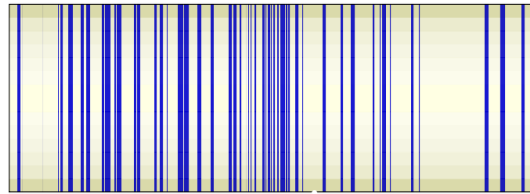
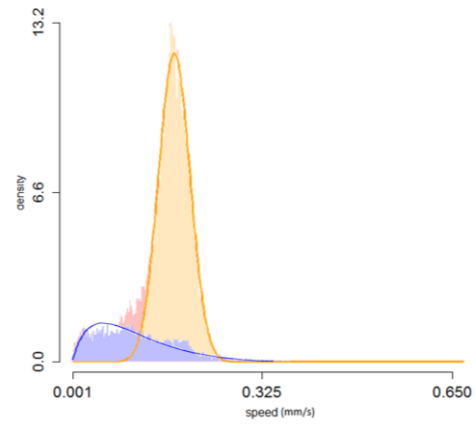
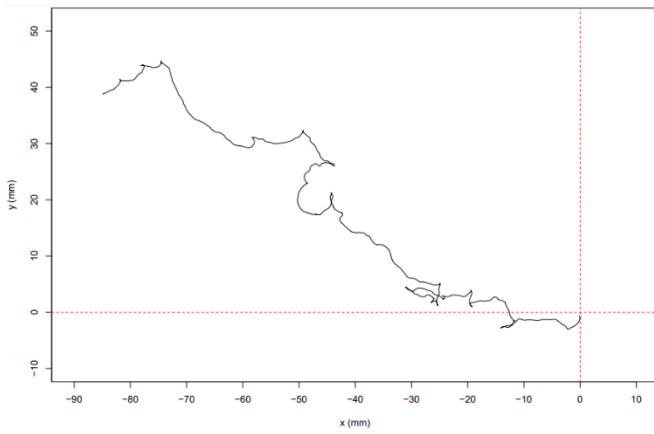


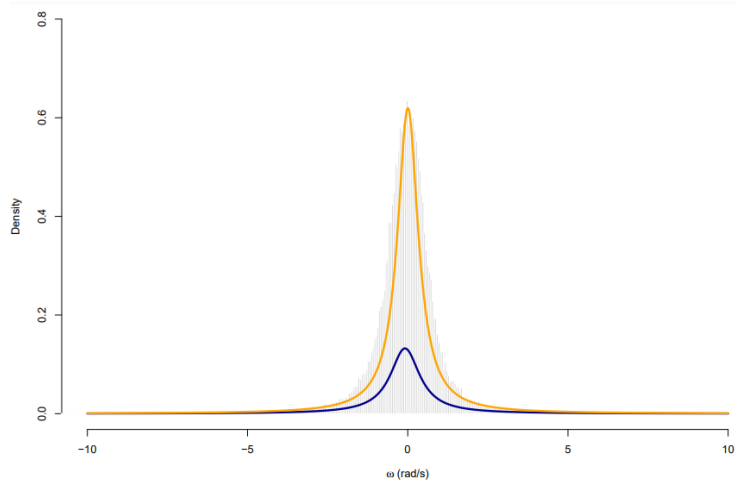
**Soca CB4856 (ind. 201406041226)**



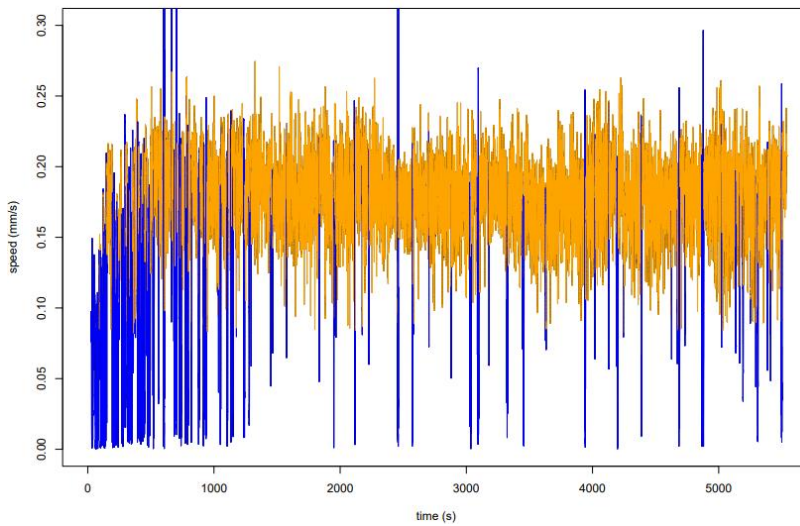
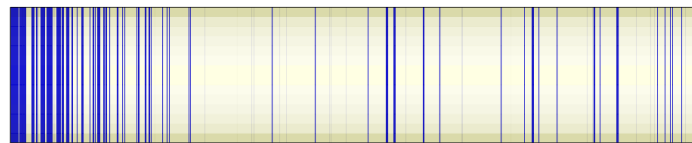
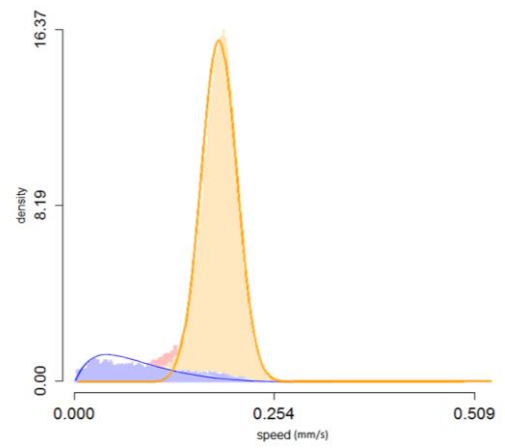
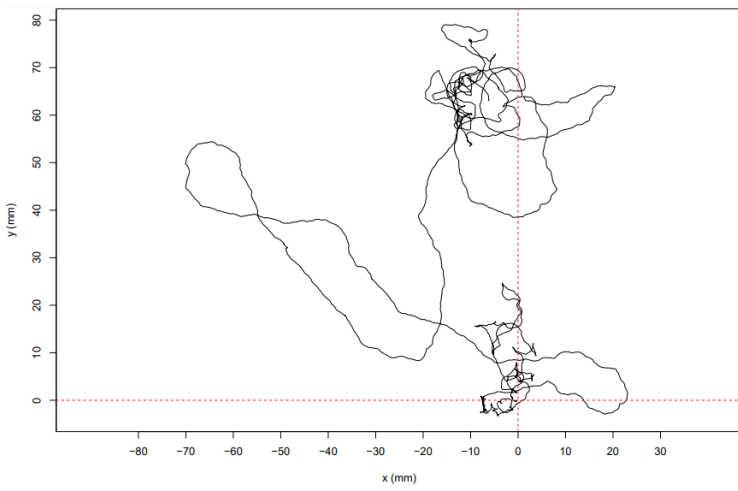


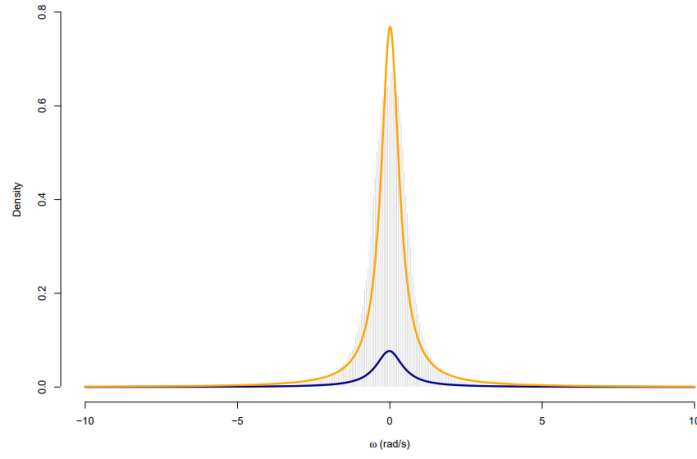
**Soca CB4856 (ind. 201406101446)**



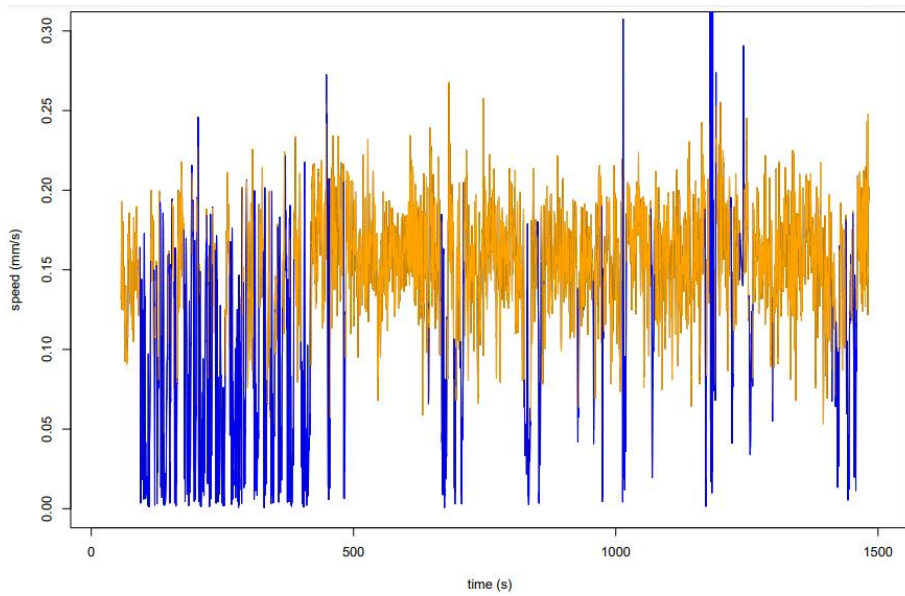
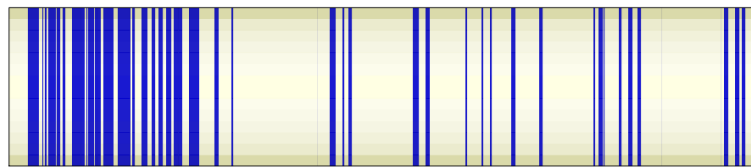
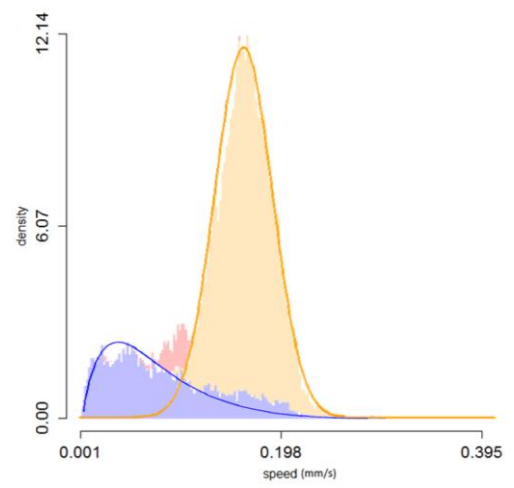
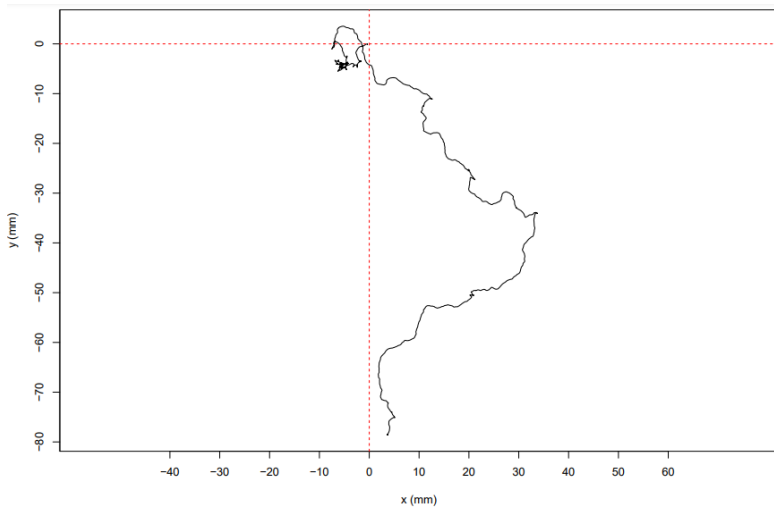


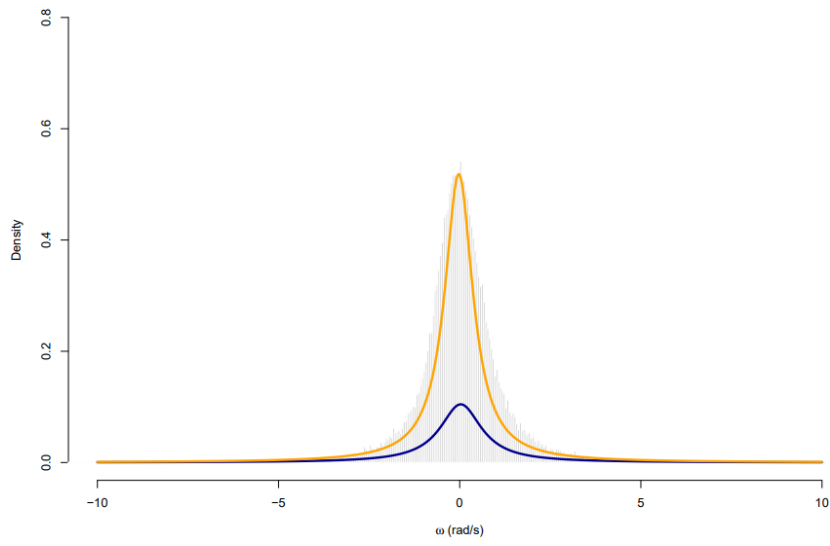
**Soca CB4856 (ind. 201406041613)**





**Soca CB4856 (ind. 201406051218)**





**Paràmetres de les equacions de Langevin per als 8 individus restants d'ambdues soques**

	Paràmetre	Valor	Unitats	Mètode
Soca N2 (ind. 201405091556)	Velocitat mitjana ( $v_s$ )	0.18677540	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02977351	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.01003404	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.07906678	rad	SD (turning angles)
Soca N2 (ind. 201405091756)	Velocitat mitjana ( $v_s$ )	0.17222604	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02817032	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.01011903	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.08445762	rad	SD (turning angles)
Soca N2 (ind. 201405101638)	Velocitat mitjana ( $v_s$ )	0.14966482	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02470945	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.005400108	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.09892767	rad	SD (turning angles)
Soca N2 (ind. 201405120011)	Velocitat mitjana ( $v_s$ )	0.14413308	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02349146	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.006277518	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.09494339	rad	SD (turning angles)
Soca CB4856 (ind. 201406041226)	Velocitat mitjana ( $v_s$ )	0.16247310	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02183764	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.003155409	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.1347543	rad	SD (turning angles)
Soca CB4856 (ind. 201406041441)	Velocitat mitjana ( $v_s$ )	0.17611183	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02712774	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.0007238411	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )



Soca CB4856 (ind. 201406041613)	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.1421975	rad	SD (turning angles)
	Velocitat mitjana ( $v_s$ )	0.18035635	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02261355	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.001713469	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
Soca CB4856 (ind. 201406051218)	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.07689212	rad	SD (turning angles)
	Velocitat mitjana ( $v_s$ )	0.15888403	mm/s	Mean (speed)
	Desviació típica de la velocitat mitjana ( $\sigma_h$ )	0.02760091	mm/s	SD (speed)
	Força de torsió ( $\beta$ )	0.004028565	rad·mm/s <sup>2</sup>	Mean (speed)*Mean( $\omega$ )
	Desviació típica de l'angle de gir ( $\sigma_\phi$ )	0.1379353	rad	SD (turning angles)

### **ANNEX III: Codi generat**

Els arxius de codi i els *pipelines* generats al llarg del present projecte es poden consultar a través del repositori GitLab identificat com TFM\_EricAlegretOlivé. Tal i com s'ha dut a terme al llarg de la metodologia plantejada, el repositori inclou dues parts diferenciades corresponents als dos blocs d'estudi proposats.