



VNIVERSITAT  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria **ETSE-UV**



PRINCIPE FELIPE  
CENTRO DE INVESTIGACION

# Alineamientos de genoma completo: su uso en la identificación taxonómica de aislados bacterianos y especiación de cepas del género *Bifidobacterium*

**TRABAJO DE FIN DE MÁSTER EN BIOINFORMÁTICA**

**Autor:** Antonio Bahilo Gómez

**Tutores:** Alfonso Benítez Páez y Rosario Gil García

# Índice

## ➤ Introducción

- Características del género *Bifidobacterium*
- Diversidad dentro de *Bifidobacterium*
- Análisis de la diversidad genética
- Análisis de genomas completos

## ➤ Objetivos

## ➤ Diseño metodológico

## ➤ Herramienta FastANI

## ➤ Resultados y discusión

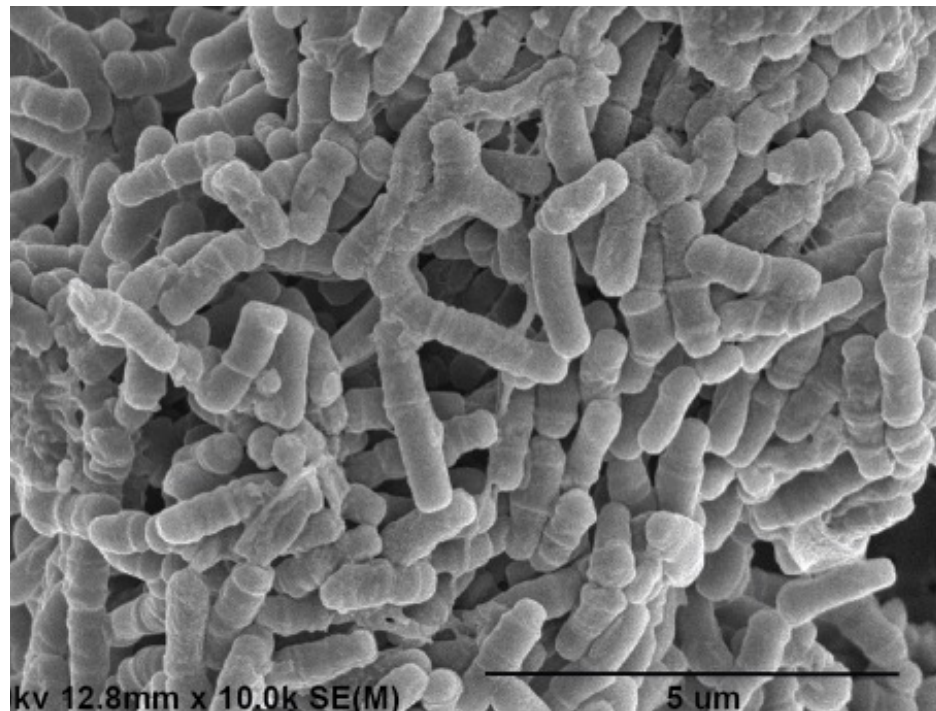
## ➤ Conclusión

## ➤ Agradecimientos

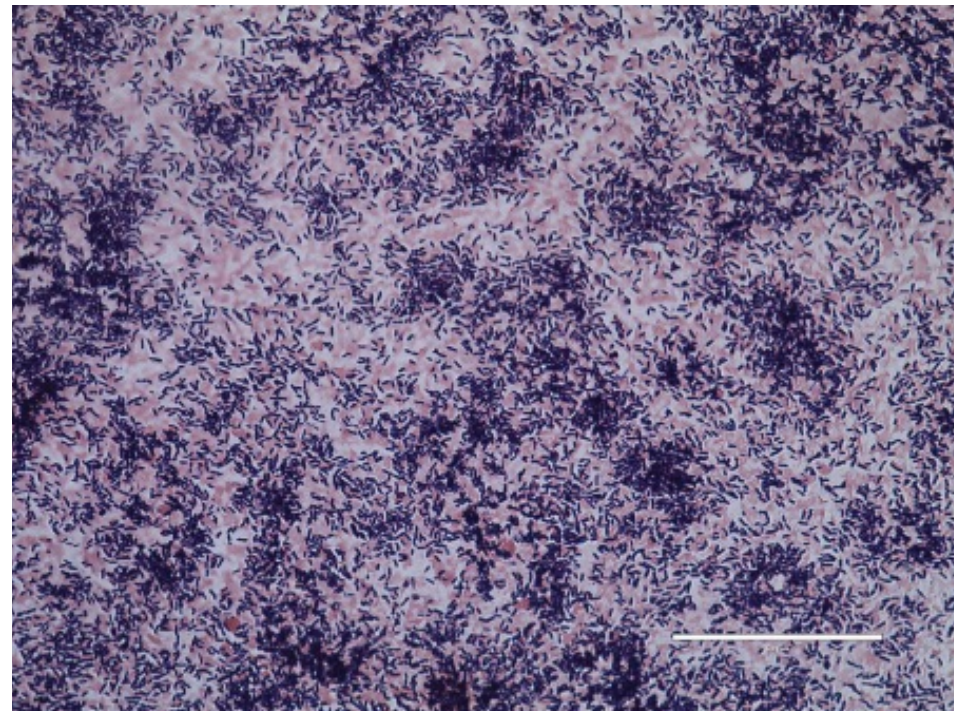


# Características del género *Bifidobacterium*

*B. infantis* (Henrick et al; 2019). DOI: [10.1038/s41390-019-0533-2](https://doi.org/10.1038/s41390-019-0533-2)

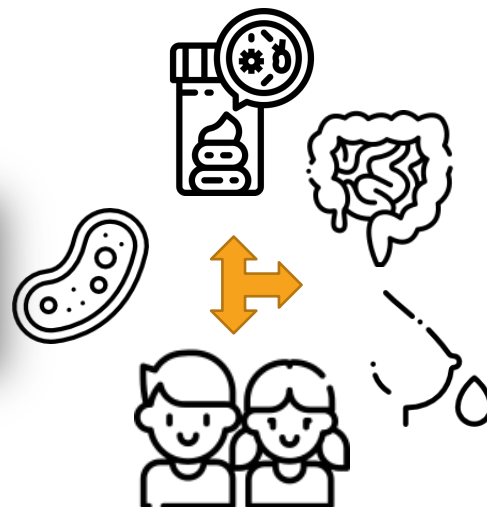


*B. infantis* (Henrick et al; 2019). DOI: [10.1038/s41390-019-0533-2](https://doi.org/10.1038/s41390-019-0533-2)



## The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota

Christian Milani <sup>1</sup>, Sabrina Duranti <sup>1</sup>, Francesca Bottacini <sup>2</sup>, Eoghan Casey <sup>2</sup>,



## The human gut microbiota during the initial stages of life: insights from bifidobacteria

Francesca Turroni <sup>1</sup>, Christian Milani <sup>2</sup>, Marco Ventura <sup>2</sup>, Douwe van Sinderen <sup>3</sup>

## Breast-Milk Microbiota Linked to Celiac Disease Development in Children: A Pilot Study From the PreventCD Cohort

Alfonso Benítez-Páez <sup>1</sup>, Marta Olivares <sup>1</sup>, Hania Szajewska <sup>2</sup>, Małgorzata Pieścik-Lech <sup>2</sup>,

# Diversidad dentro de *Bifidobacterium*

## Genomic insights into bifidobacteria

Ju-Hoon Lee <sup>1</sup>, Daniel J O'Sullivan

Species no. <sup>a</sup>	Name <sup>b</sup>	Subspecies	Origin <sup>c</sup>
1	<i>B. adolescentis</i>		Intestine of adult
2	<i>B. angulatum</i>		Human feces
3	<i>B. animalis</i>	<i>B. animalis</i> subsp. <i>animalis</i> <i>B. animalis</i> subsp. <i>lactis</i>	Animal feces Yogurt
4	<i>B. asteroides</i>		Intestine of honeybee
5	<i>B. bifidum</i>		Infant feces
6	<i>B. bombi</i>		Intestine of bumblebee
7	<i>B. boum</i>		Rumen of cattle
8	<i>B. breve</i>		Intestine of infant
9	<i>B. catenulatum</i>		Intestine of adult
10	<i>B. choerinum</i>		Porcine feces
11	<i>B. coryneforme</i>		Intestine of honeybee
12	<i>B. crudilactis</i>		Raw milk
13	<i>B. cuniculi</i>		Feces of rabbit
14	<i>B. dentium</i>		Human dental caries
15	<i>B. gallicum</i>		Human feces
16	<i>B. gallinarum</i>		Chicken cecum
17	<i>B. indicum</i>		Intestine of honeybee
18	<i>B. longum</i>	<i>B. longum</i> subsp. <i>infantis</i> <i>B. longum</i> subsp. <i>longum</i> <i>B. longum</i> subsp. <i>suis</i>	Intestine of infant Intestine of adult Porcine feces
19	<i>B. magnum</i>		Rabbit feces
20	<i>B. merycicum</i>		Bovine rumen
21	<i>B. minimum</i>		Sewage
22	<i>B. pseudocatenulatum</i>		Infant feces
23	<i>B. pseudolongum</i>	<i>B. pseudolongum</i> subsp. <i>globosum</i> <i>B. pseudolongum</i> subsp. <i>pseudolongum</i>	Bovine rumen Porcine feces
24	<i>B. psychraerophilum</i>		Porcine feces
25	<i>B. pullorum</i>		Chicken feces
26	<i>B. ruminantium</i>		Bovine rumen
27	<i>B. saeculare</i>		Rabbit feces
28	<i>B. scardovii</i>		Human blood
29	<i>B. subtile</i>		Sewage
30	<i>B. thermophilum</i>		Porcine feces
31	<i>B. thermacidophilum</i>	<i>B. thermacidophilum</i> subsp. <i>porcinum</i> <i>B. thermacidophilum</i> subsp. <i>thermacidophilum</i>	Porcine feces Sewage

## Beneficios para la salud humana

### Acidified milk formula supplemented with bifidobacterium lactis: impact on infant diarrhea in residential care settings

Jean-Pierre Chouraqui <sup>1</sup>, Louis-Dominique Van Egroo, Marie-Claire Fichot

### Effects of yogurt and bifidobacteria supplementation on the colonic microbiota in lactose-intolerant subjects

T He <sup>1</sup>, M G Priebe, Y Zhong, C Huang, H J M Harmsen, G C Raangs, J-M Antoine, G W Welling,

### Cholesterol-lowering effect of probiotic yogurt in comparison with ordinary yogurt in mildly to moderately hypercholesterolemic subjects

Asal Ataie-Jafari <sup>1</sup>, Bagher Larijani, Hamid Alavi Majd, Farideh Tahbaz

### Immunomodulating potential of supplementation with probiotics: a dose-response study in healthy young adults

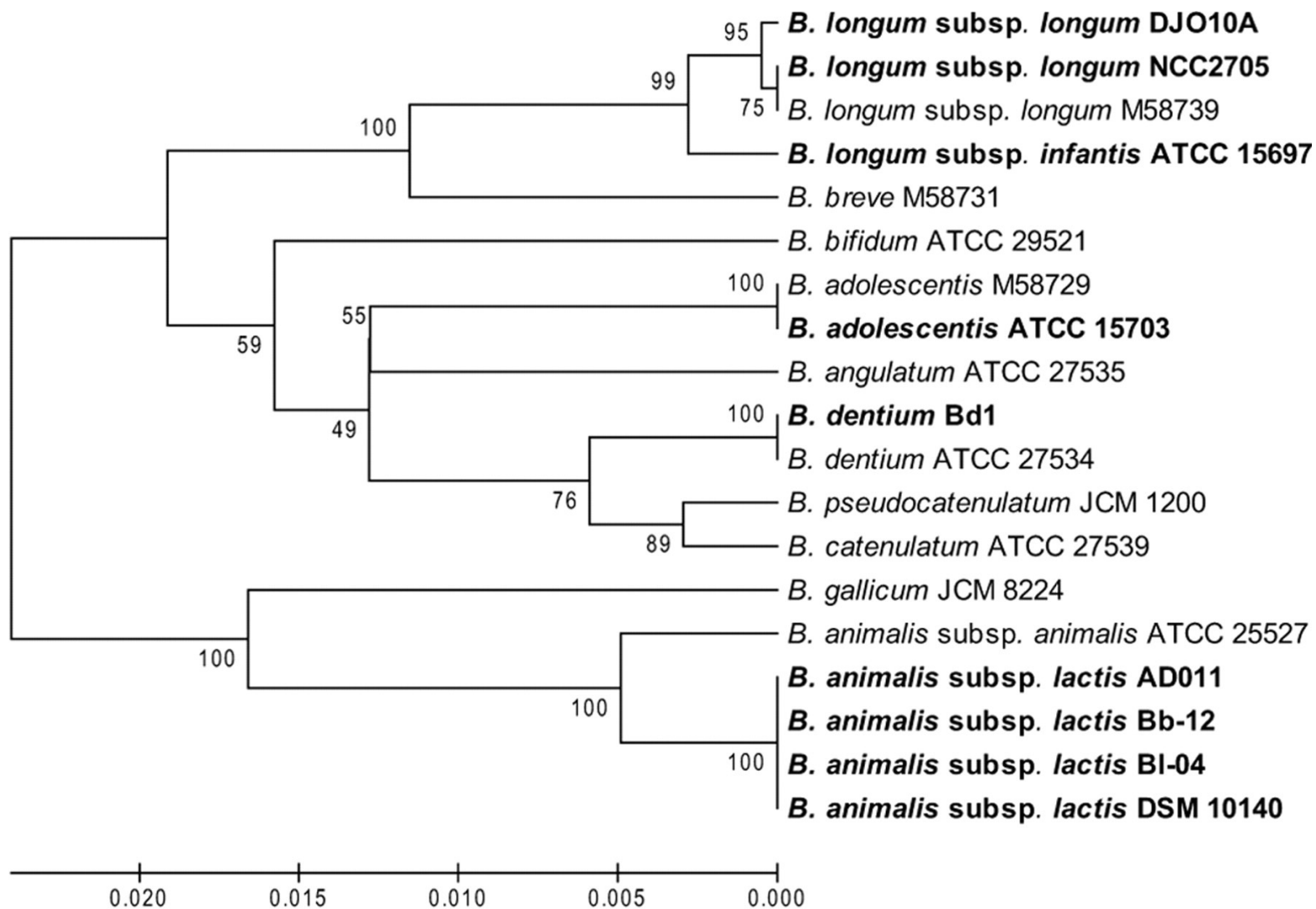
Hanne Risager Christensen <sup>1</sup>, Charlotte Nexmann Larsen, Pernille Kaestel, Lisbeth Buus Rosholm,

### Synbiotic intervention of Bifidobacterium lactis and resistant starch protects against colorectal cancer development in rats

Richard K Le Leu <sup>1</sup>, Ying Hu, Ian L Brown, Richard J Woodman, Graeme P Young



# Análisis de la diversidad biológica



Extraído de (Lee & O'Sullivan; 2010). DOI: [10.1128/MMBR.00004-10](https://doi.org/10.1128/MMBR.00004-10)

## Cronología de las técnicas de análisis filogenético

DHH

1961

ARN  
ribosómico  
16S

1977

MLST de genes  
codificantes

1998

ANI (identidad  
nucleotídica  
media)

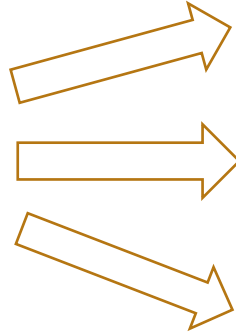
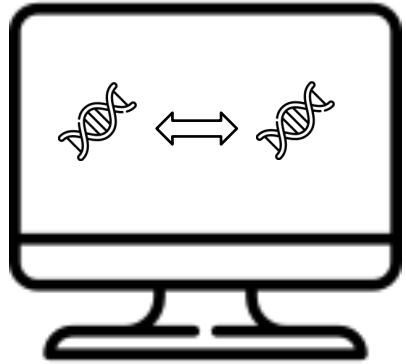
2005

**Determinación  
de especie**

- DHH >70%
- ARNr 16S > 97%
- ANI > 95%

Extraído de (Hugenholtz et al; 2021). DOI: [10.1038/s41396-021-00941-x](https://doi.org/10.1038/s41396-021-00941-x)

# Análisis de genomas completos



**ANIm**

**ANIb**

**FastANI**

**A large-scale evaluation of algorithms to calculate average nucleotide identity**

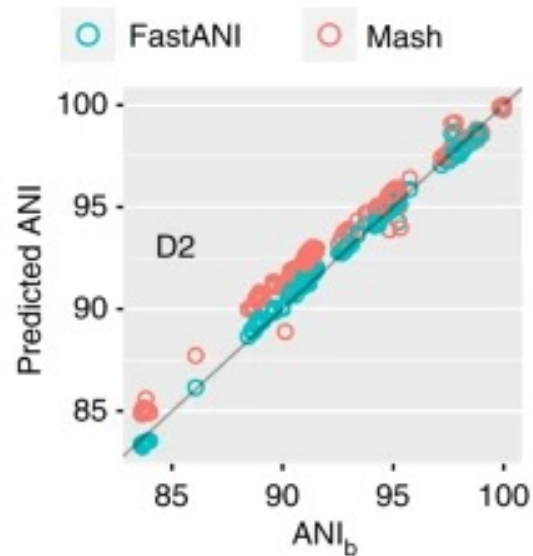
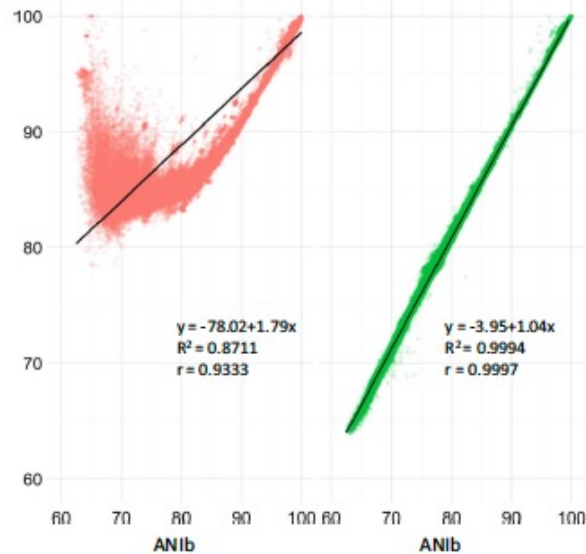
[Seok-Hwan Yoon](#)<sup>1,2</sup>, [Sung-Min Ha](#)<sup>1,2</sup>, [Jeongmin Lim](#)<sup>2</sup>, [Soonjae Kwon](#)<sup>2</sup>, [Jongsik Chun](#)<sup>3,4</sup>

**High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries**

[Chirag Jain](#), [Luis M. Rodriguez-R](#), [Adam M. Phillippy](#), [Konstantinos T. Konstantinidis](#) & [Srinivas Aluru](#)

**(A) ANIm**

**(B) OrthoANIb**



## **ANIm VS ANIb**

- ANIm comete errores en rangos <90% de ANI.
- ANIm es 4,7x más rápido que ANIb.

## **FastANI VS ANIb**

- FastANI misma precisión que ANIb.
- FastANI 782x más rápido que ANIb.
- FastANI permite comparar genomas incompletos.

## Objetivos

- El objetivo general es realizar un **análisis comparativo de más de 800 genomas de cepas pertenecientes al género *Bifidobacterium*** con el estándar **ANI**. Los objetivos específicos son:
  - Establecer **relaciones filogenéticas más robustas** entre las diferentes especies de bifidobacterias con la herramienta **FastANI**.
  - **Detectar inconsistencias** generadas a partir de **anotaciones taxonómicas erróneas** entre los genomas descargados de la base de datos BV-BCR.
  - **Establecer** distribución de **valores ANI intra- e interespecie** en el grupo de datos analizado.

# Diseño metodológico

CBV-BRC 3.28.9 ORGANISMS SEARCHES TOOLS & SERVICES All Data Types Find a gene, genome, microarray, etc

Genome List View  
GENOMES: ( 822 Genomes )

Overview Genomes Sequences AMR Phenotypes Features Proteins Specialty Genes Pathways Subsystems

DOWNLOAD KEYWORDS ADV Search FILTERS APPLY

<input type="checkbox"/>	Genome Name	Strain	GenBank Accessions	Size	CDS	Collection Year	Isolation Country	Host Common Name
<input type="checkbox"/>	Bifidobacterium longum strain Bifidobact	Bifidobacterium_longum_BgEED06	CABFLL000000000	2426212	2164			Human
<input type="checkbox"/>	Bifidobacterium longum subsp. longum s	67B	JAERWH000000000	2386390	2139	2020	Italy	Human
<input type="checkbox"/>	Bifidobacterium longum subsp. longum s	108B	JAERWG000000000	2305647	1991	2020	Italy	Human
<input type="checkbox"/>	Bifidobacterium longum subsp. suis strai	209B	JAERWE000000000	2300337	2041	2020	Italy	Human



FastANI



Interpretación  
resultados

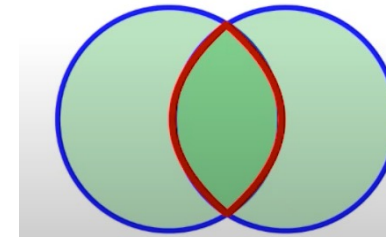
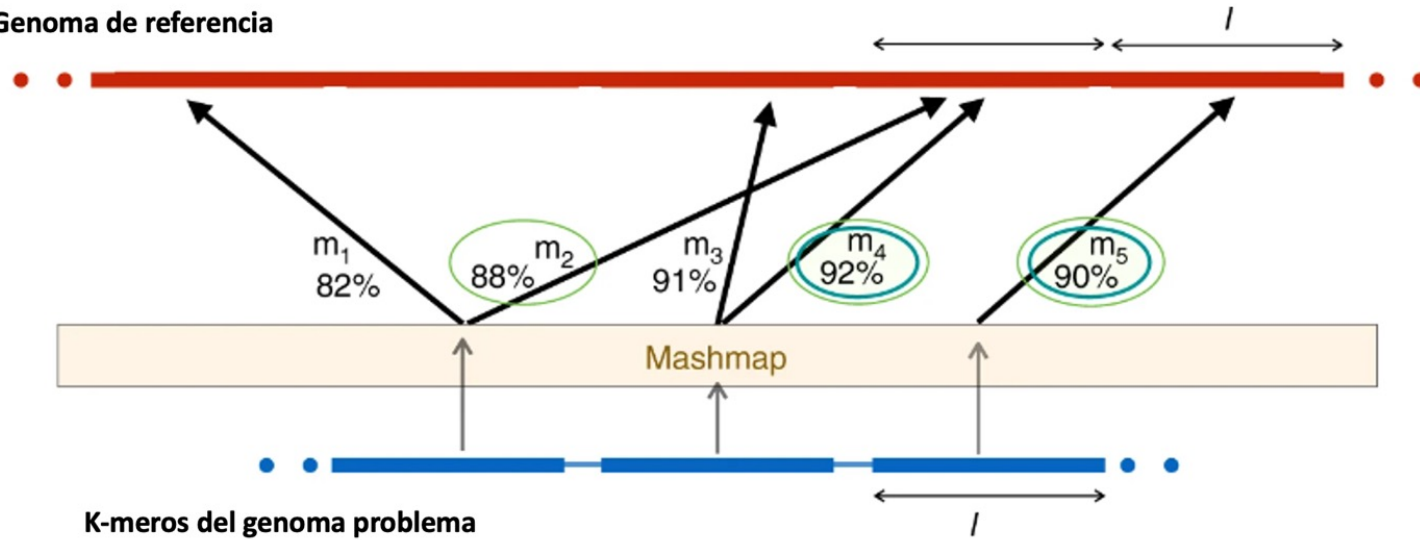


Filtrado



# Algoritmo FastANI

Genoma de referencia

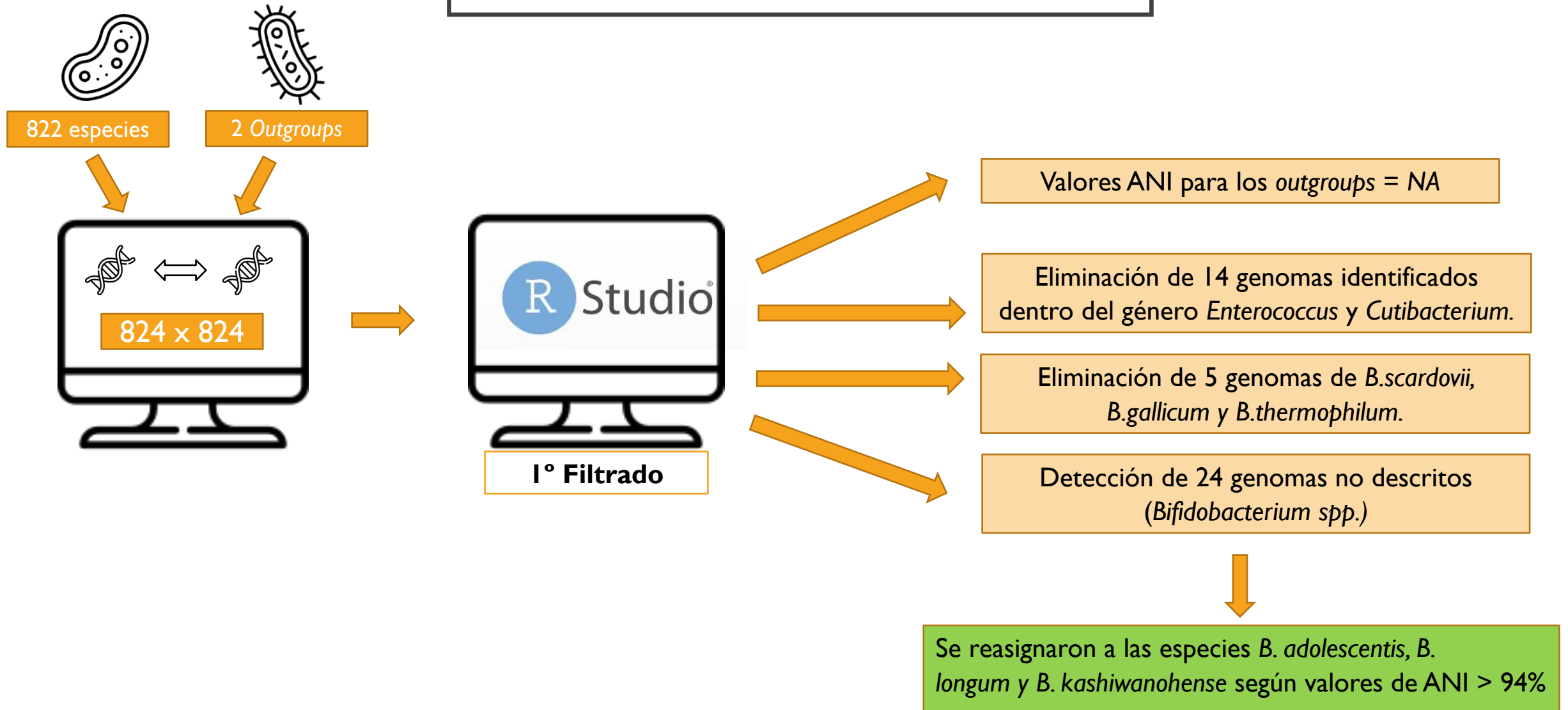


$$\frac{|A \cap B|}{|A \cup B|} = J$$

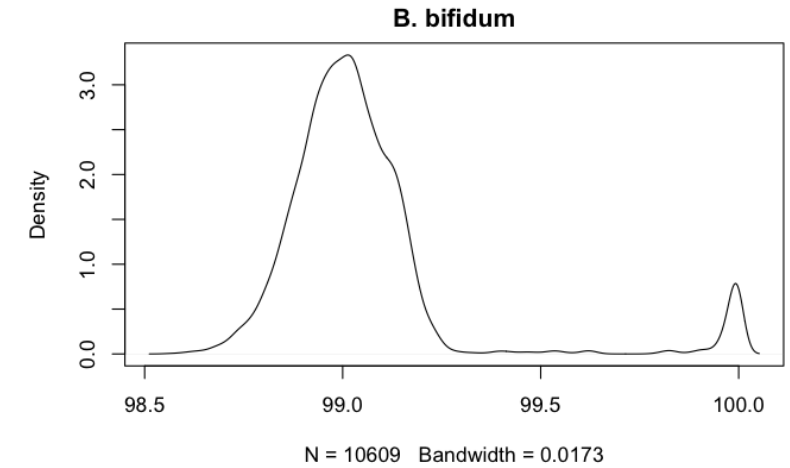
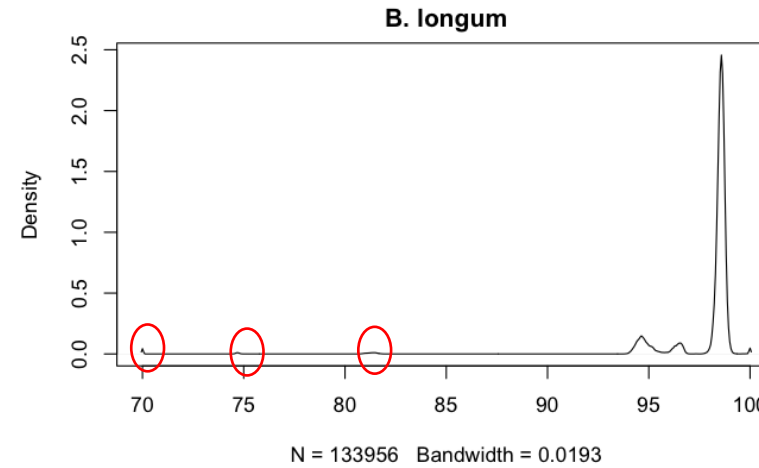
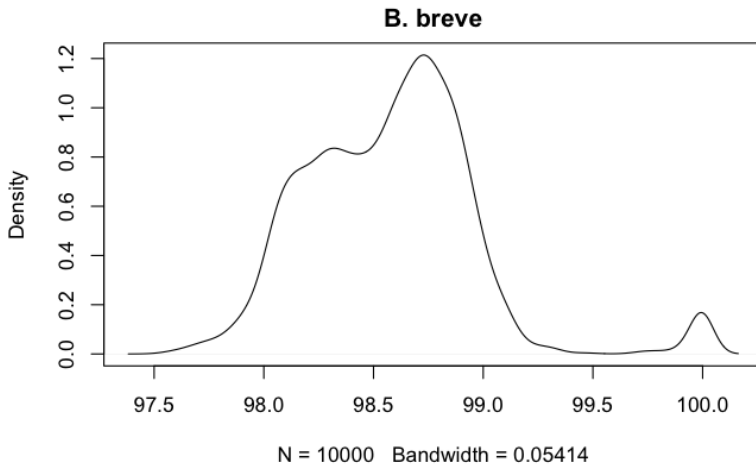
$J$  is the Jaccard coefficient

Extraído de (Jain et al., 2018). DOI: [10.1089/cmb.2018.0036](https://doi.org/10.1089/cmb.2018.0036)

# Resultados y discusión



# Resultados y discusión



## Relaciones Intraespecie

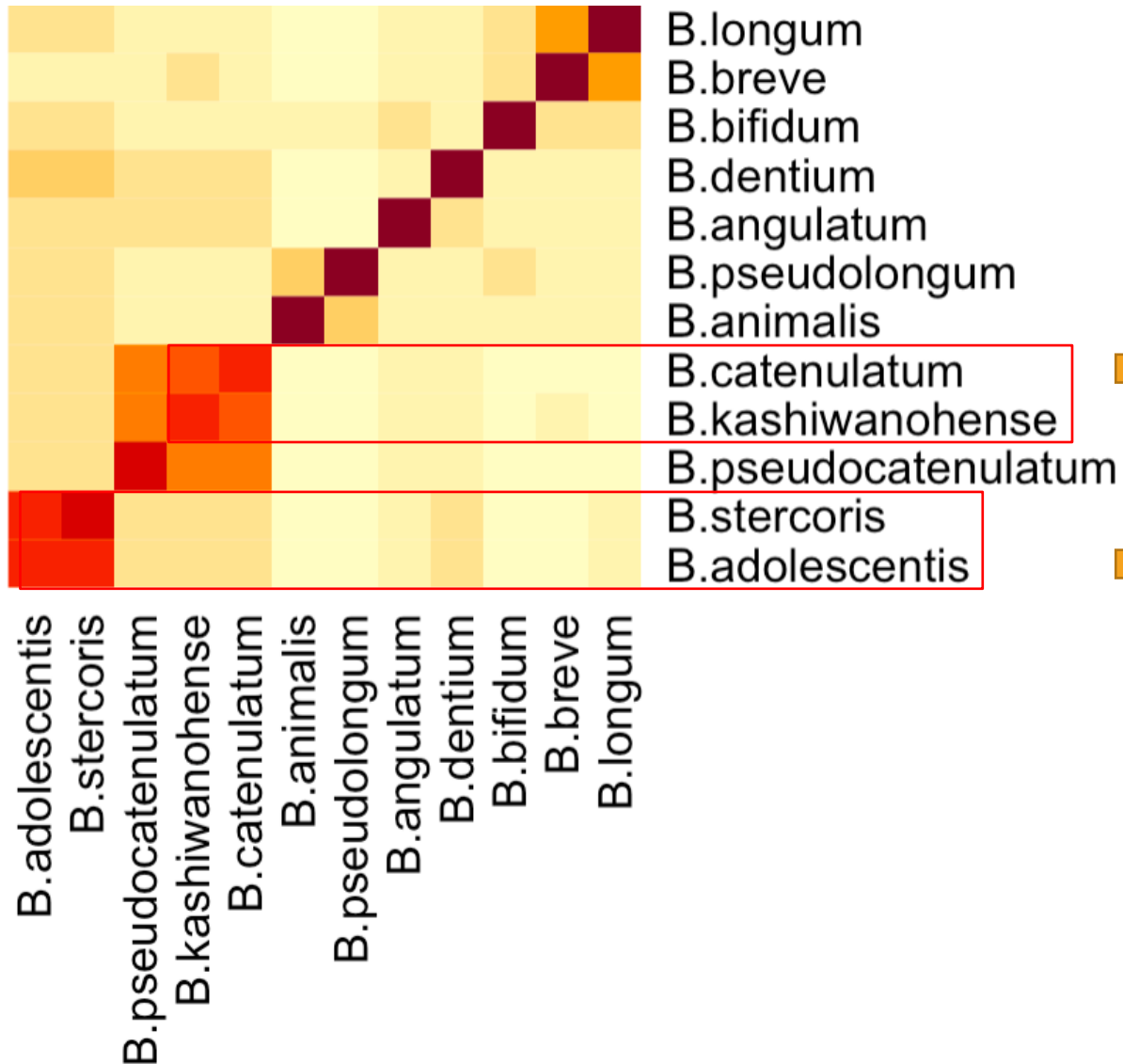
	<i>B.breve</i>	<i>B.kashiwanohense</i>	<i>B.animalis</i>	<i>B.bifidum</i>	<i>B.adolescentis</i>	<i>B.longum</i>
Tamaño_N	10000	16	1444	10609	7225	133956
Media_ANI_intra	98,57	96,31	99,60	99,05	98,22	97,89
Desviación	0,39	2,40	1,19	0,24	0,36	2,59
CI 95%	98,56-98,57	95,13-97,48	99,53-99,65	99,04-99,05	98,21-98,22	97,87-97,90
Mínimo	97,54	93,99	95,95	98,56	97,15	70,00
Mediana	98,59	95,04	99,99	99,01	98,22	98,53
Cuartil Q1	98,29	94,53	99,99	98,93	98,01	98,36
Cuartil Q3	98,80	98,03	100,00	99,09	98,39	98,66
	<i>B.pseudocatenulatum</i>	<i>B.pseudolongum</i>	<i>B.angularum</i>	<i>B.catenulatum</i>	<i>B.dentium</i>	<i>B.stercoris</i>
Tamaño_N	4900,00	4,00	9,00	49,00	529,00	1,00
Media_ANI_intra	98,47	98,97	99,37	98,22	99,01	100,00
Desviación	0,42	1,19	0,74	1,34	0,45	NA
CI 95%	98,45-98,48	97,79-100	98,88-99,85	97,83-98,59	98,97-99,05	NA-NA
Mínimo	96,73	97,92	98,56	96,01	98,14	100,00
Mediana	98,50	98,97	99,99	98,40	98,82	100,00
Cuartil Q1	98,26	97,94	98,60	98,13	98,70	100,00
Cuartil Q3	98,69	100,00	100,00	98,61	99,17	100,00

## Resultados y discusión

### Relaciones Interespecie

	<i>B.breve</i>	<i>B.kashiwanohense</i>	<i>B.animalis</i>	<i>B.bifidum</i>	<i>B.adolescentis</i>	<i>B.longum</i>
Tamaño_N	70200,00	3192,00	29032,00	71997,00	60945,00	159576,00
Media_ANI_inter	83,15	81,16	77,60	80,13	80,54	81,42
Desviación	3,60	4,00	0,44	1,01	1,68	3,02
CI 95%	83,12-83,18	81,02-81,29	77,59-77,60	80,11-80,13	80,52-80,54	81,40-81,43
Mínimo	70,00	74,81	70,00	70,00	70,00	70,00
Mediana	85,97	79,57	77,58	80,50	80,17	80,50
Cuartil Q1	79,75	79,26	77,47	79,77	79,81	79,60
Cuartil Q3	86,53	80,61	77,68	80,74	80,43	81,19
	<i>B.pseudocatenulatum</i>	<i>B.pseudolongum</i>	<i>B.angulatum</i>	<i>B.catenulatum</i>	<i>B.dentium</i>	<i>B.stercoris</i>
Tamaño_N	51240,00	1600,00	2397,00	5565,00	17917,00	801,00
Media_ANI_inter	80,18	78,11	79,95	80,93	79,56	82,38
Desviación	2,26	0,81	0,97	3,97	1,54	5,63
CI 95%	80,16-80,19	78,07-78,15	79,90-79,98	80,82-81,03	79,53-79,58	81,98-82,76
Mínimo	70,00	70,00	70,00	70,00	70,00	74,71
Mediana	79,55	77,95	79,82	79,27	78,96	80,20
Cuartil Q1	79,28	77,75	79,52	79,03	78,76	79,89
Cuartil Q3	79,85	78,31	80,09	80,04	79,18	80,56

## Resultados y discusión



**Comparative genomic analysis revealed genetic divergence between *Bifidobacterium catenulatum* subspecies present in infant versus adult guts**

Jiaqi Liu <sup># 1</sup>, Weicheng Li <sup># 1</sup>, Caiqing Yao <sup>1</sup>, Jie Yu <sup>1</sup>, Heping Zhang <sup>2</sup>

*B. kashiwanohense* y *B. catenulatum* tenían un valor de ANI interespecie > 95%

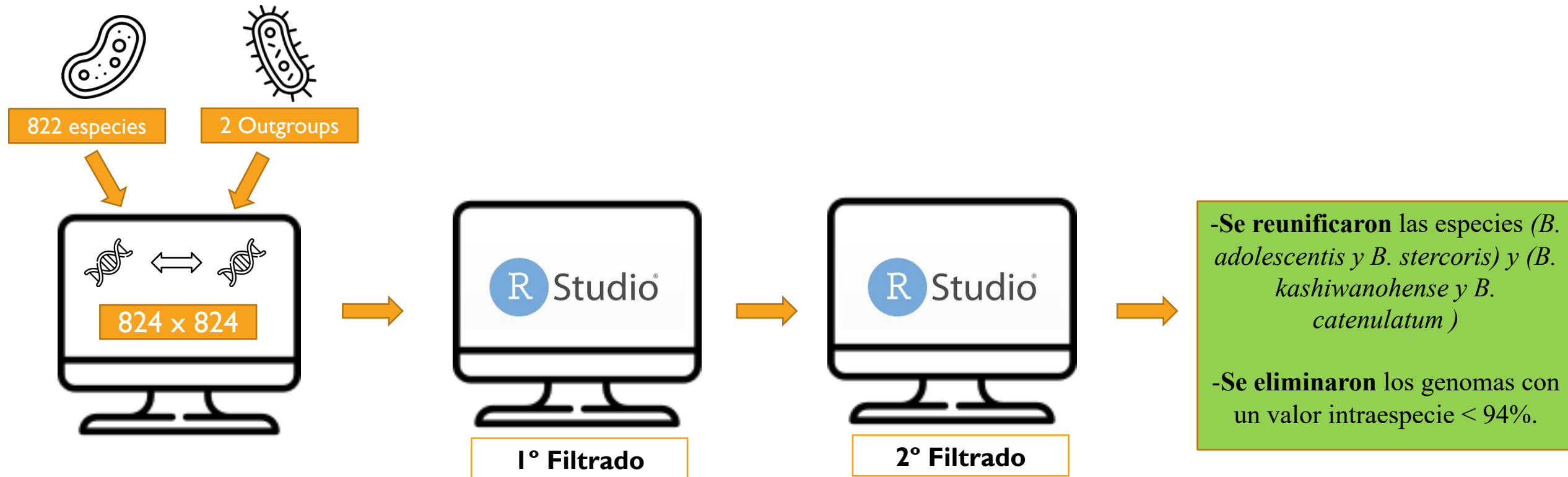
*B. adolescentis* y *B. stercoris* tenían un valor de ANI interespecie > 95%.

**Reclassification of *Bifidobacterium stercoris* Kim et al. 2010 as a later heterotypic synonym of *Bifidobacterium adolescentis***

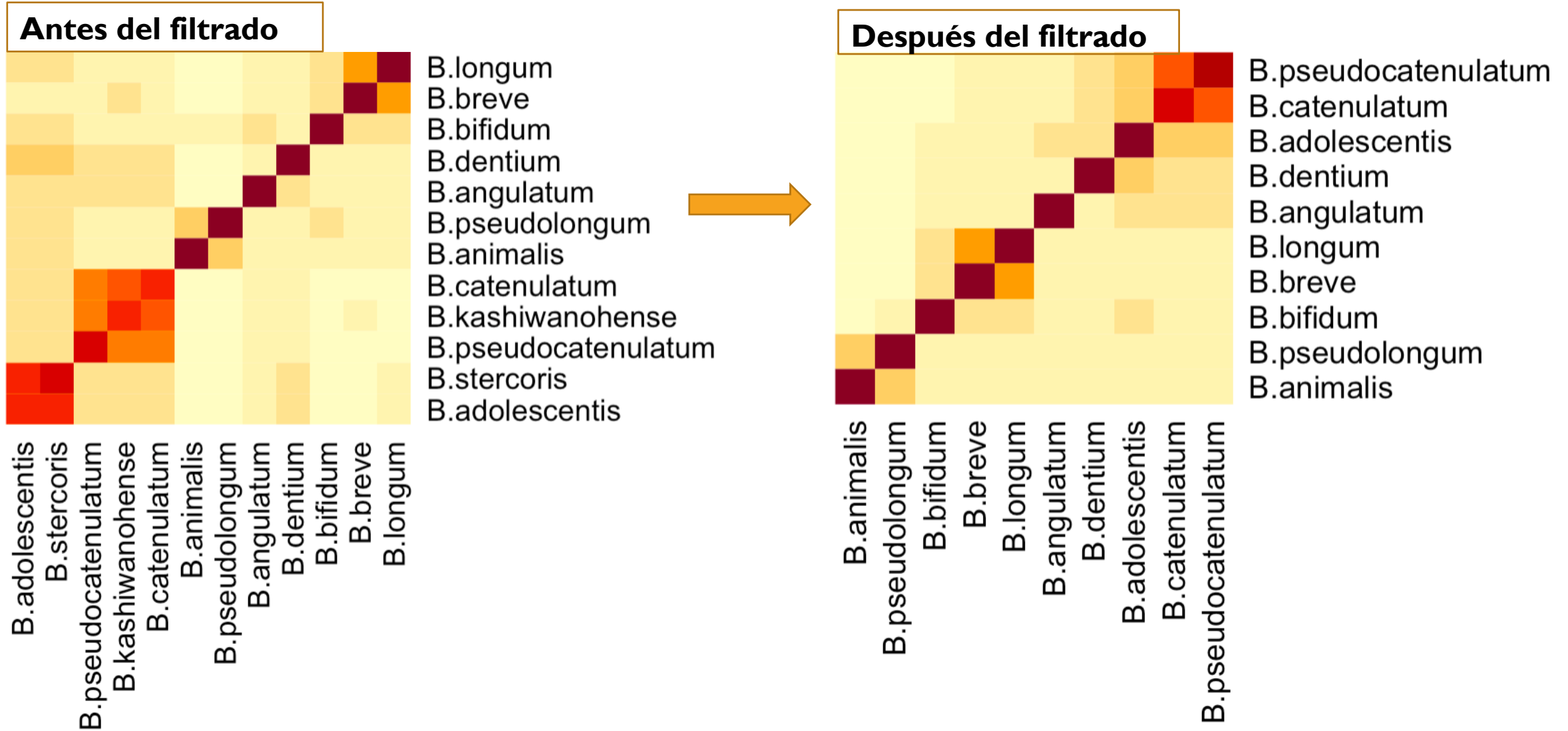
J Killer <sup>1 2</sup>, I Sedláček <sup>3</sup>, V Rada <sup>1</sup>, J Havlík <sup>1</sup>, J Kopečný <sup>2</sup>



## Resultados y discusión



## Resultados y discusión



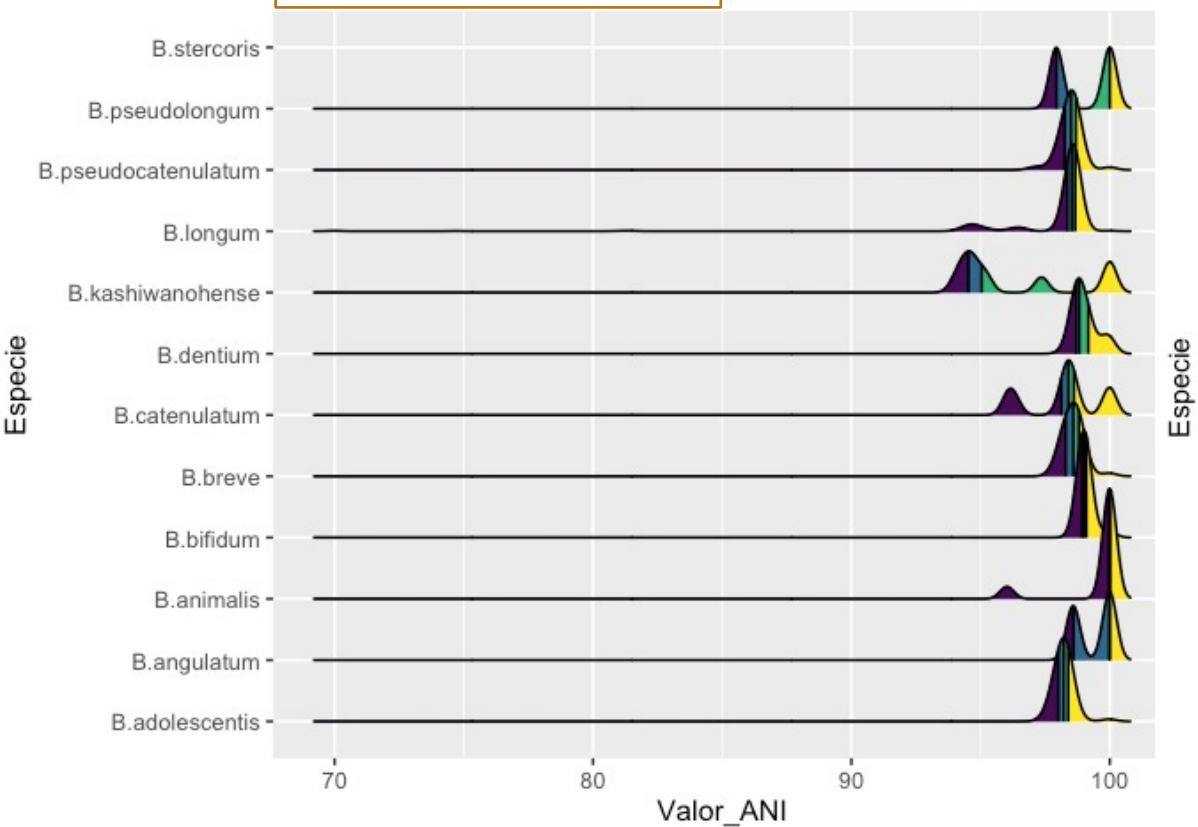
## Resultados y discusión

### Relaciones Interespecie

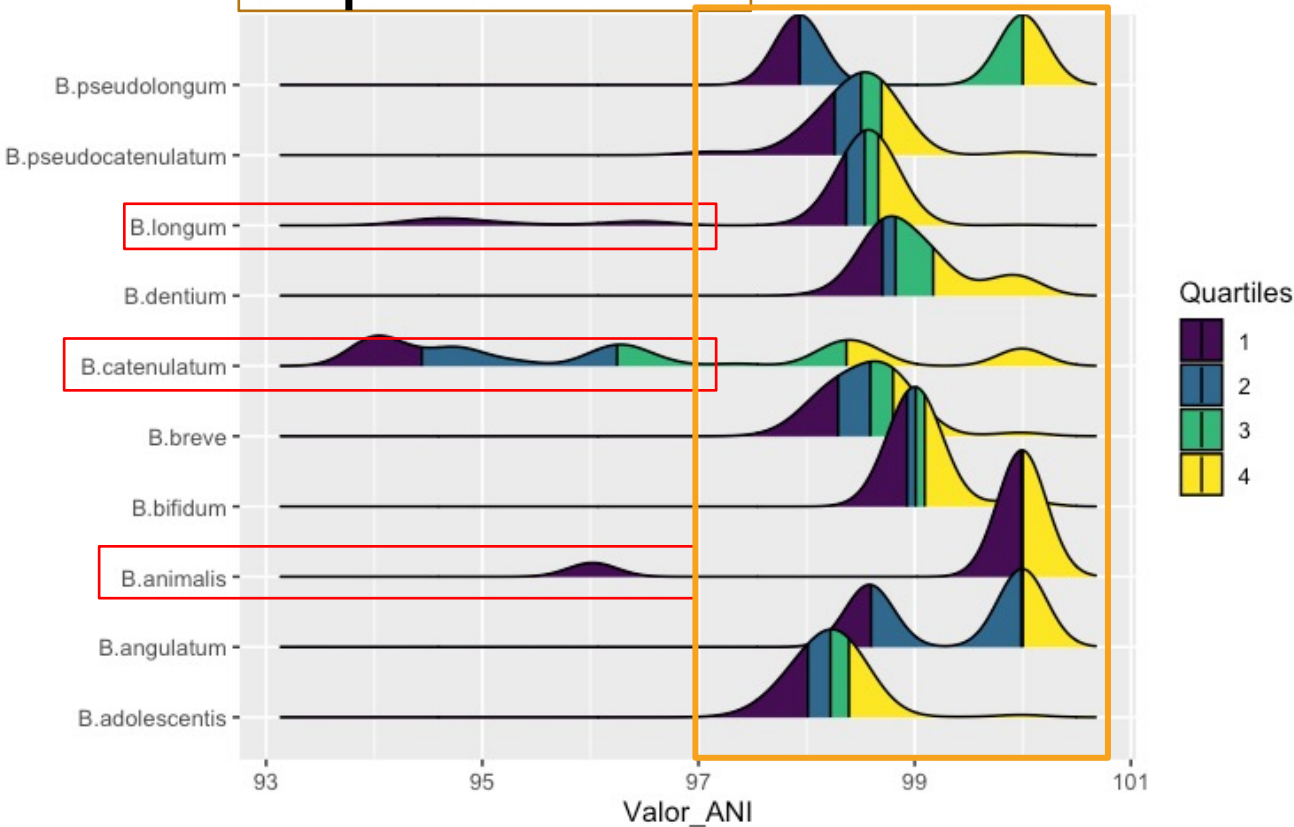
	<i>B.breve</i>	<i>B.catenulatum</i>	<i>B.animalis</i>	<i>B.bifidum</i>	<i>B.adolescentis</i>
Tamaño_N	70000,00	8679,00	28956,00	71791,00	61404,00
Media_ANI_inter	83	81	78	80	81
Desviación	3,57	3,83	0,34	0,94	1,50
CI 95%	83,14-83,20	80,85-81,01	77,60-77,61	80,13-80,14	80,51-80,53
Mínimo	76,80	77,12	76,74	77,45	77,35
Mediana	85,98	79,37	77,58	80,50	80,17
Cuartil Q1	79,75	79,08	77,47	79,78	79,81
Cuartil Q3	86,53	80,24	77,68	80,74	80,43
	<i>B.longum</i>	<i>B.pseudocatenulatum</i>	<i>B.pseudolongum</i>	<i>B.angularatum</i>	<i>B.dentium</i>
Tamaño_N	158704,00	51100,00	1596,00	2391,00	17871,00
Media_ANI_inter	81	80	78	80	80
Desviación	2,98	2,25	0,75	0,90	1,53
CI 95%	81,43-81,46	80,17-80,21	78,08-78,15	79,92-79,99	79,54-79,59
Mínimo	76,99	77,19	77,00	77,60	77,2
Mediana	80,51	79,55	77,95	79,82	78,97
Cuartil Q1	79,60	79,28	77,76	79,52	78,76
Cuartil Q3	81,20	79,85	78,31	80,09	79,18

# Resultados y discusión

Antes del filtrado



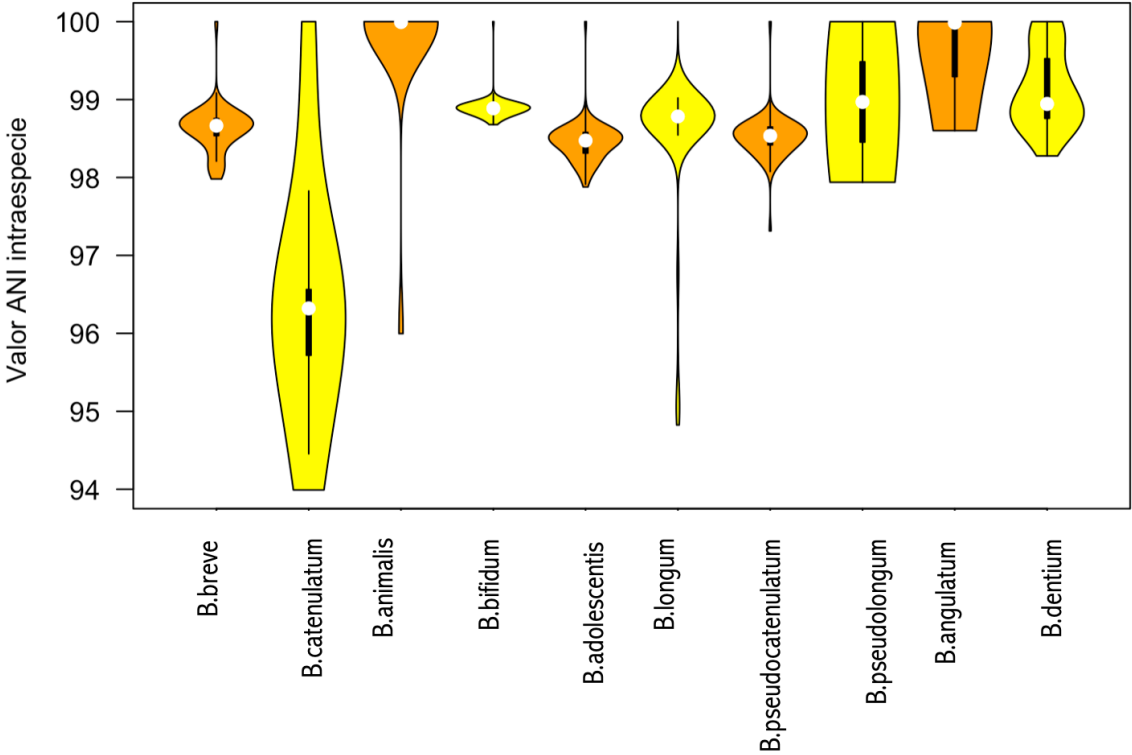
Después del filtrado



# Resultados y discusión

## Relaciones Intraespecie

	<i>B.breve</i>	<i>B.catenulatum</i>	<i>B.animalis</i>	<i>B.bifidum</i>	<i>B.adolescentis</i>
Tamaño_N	10000,00	121,00	1444,00	10609,00	7396,00
Media_ANI_intra	98,57	96,44	99,60	99,05	98,22
Desviación	0,39	2,10	1,19	0,24	0,36
CI 95%	98,56-98,57	96,06-96,81	99,53-99,65	99,04-99,05	98,21-98,22
Mínimo	97,5	94,0	95,9	98,6	97,1
Mediana	98,59	96,25	99,99	99,01	98,22
Cuartil Q1	98,29	94,44	99,99	98,93	98,01
Cuartil Q3	98,80	98,37	100,00	99,09	98,39
	<i>B.longum</i>	<i>B.pseudocatenulatum</i>	<i>B.pseudolongum</i>	<i>B.angulatum</i>	<i>B.dentium</i>
Tamaño_N	132496,00	4900,00	4,00	9,00	529,00
Media_ANI_intra	98,12	98,47	98,97	99,37	99,01
Desviación	1,19	0,42	1,19	0,74	0,45
CI 95%	98,11-98,12	98,45-98,48	97,79-100,13	98,88-99,85	98,97-99,05
Mínimo	94,1	96,7	97,9	98,6	98,1
Mediana	98,54	98,50	98,97	99,99	98,82
Cuartil Q1	98,37	98,26	97,94	98,60	98,70
Cuartil Q3	98,66	98,69	100,00	100,00	99,17





## CONCLUSIÓN

1. Debería de ser necesario **realizar siempre un filtrado** inicial **antes** de realizar cualquier **estudio genómico**.
2. El uso de la **herramienta bioinformática FastANI** ha resultado ser un excelente método de análisis para las **relaciones intraespecies**, siendo el valor **ANI obtenido en este estudio mayor a 94%** para la clasificación de genomas dentro de una misma especie, y **de 97-98% para las especies con menor diversidad** entre sus genomas.
3. La **reasignación de especies** será una **fase necesaria** para cualquier **análisis genómico** que use genomas descargados de repositorios públicos.
4. El **establecimiento de la distribución de valores ANI interespecie** en el grupo de datos analizado, ha mostrado que **el intervalo de (80-94%) podría ser útil** para distinguir la variación género específica.

# Agradecimientos

