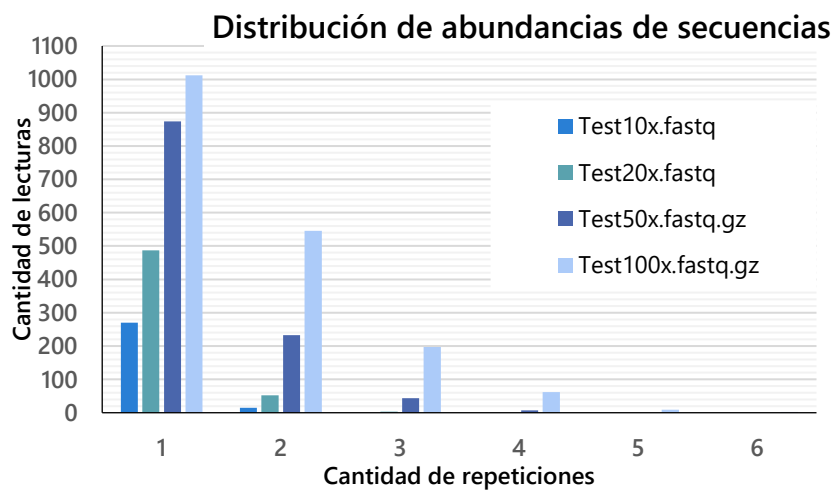


Ensamblaje de genomas

Resultados obtenidos de los Sobrelapes

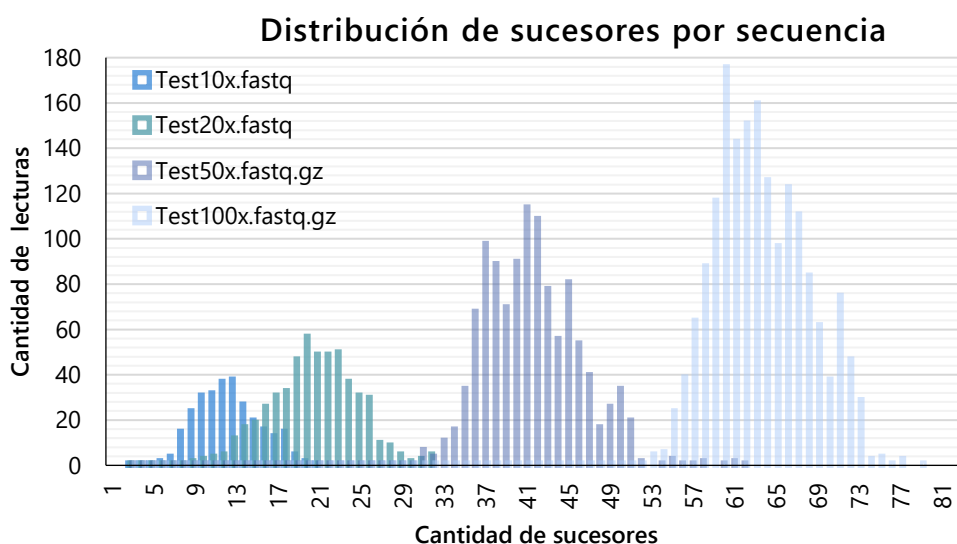
Distribución de abundancias de secuencias



Gráfica 1 Distribución de abundancias de secuencias. Los datos con lo que se hizo la gráfica se encuentran en la Tabla 3

En la Gráfica 1, es posible observar que la cantidad de secuencias encontradas depende de la profundidad de secuenciación. De igual forma, la cantidad de secuencias que se encuentran repetidas también aumenta, por ejemplo, en el archivo [test10x.fastq](#) únicamente hay 15 secuencias que se encuentran repetidas dos veces, mientras que en el caso del archivo [test100x.fastq.gz](#) hay 546 secuencias. Así mismo, el archivo [test100x.fastq.gz](#) es el único en el que se encuentran secuencias que están repetidas hasta 6 veces.

Distribución de sucesores por secuencia



Gráfica 2 Distribución de sucesores por secuencia. Los datos con los que se hizo la gráfica se encuentran en la Tabla 4

De acuerdo con la Gráfica 2, se encontró que a medida que aumenta la profundidad de la secuenciación, es mas probable encontrar sucesores para cada una de las secuencias, por ejemplo, el rango de sucesores del archivo **test10x.fastq** es de [2,20], mientras que en el archivo **test20x.fastq** el rango aumenta a [2,31] y así hasta llegar al archivo **test100x.fastq.gz** que tiene un rango de sucesores de [2,82].

Tiempo de la construcción de los sobrelapes

Tabla 1 Tiempo (ms) en el que se demora construyendo la tabla de sobrelapes.

Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
226	404	1287	2942

Debido a que la cantidad de secuencias únicas aumenta con la profundidad de la secuenciación, tal como se muestra en la Gráfica 1, el tiempo en el que se demora almacenando la información de los posibles sobrelapes de cada una de las secuencias es mayor en el archivo que tiene más secuencias (**test100x.fastq.gz**).

Alineamiento

Tabla 2 Resultados del alineamiento de secuencias de cada uno de los archivos

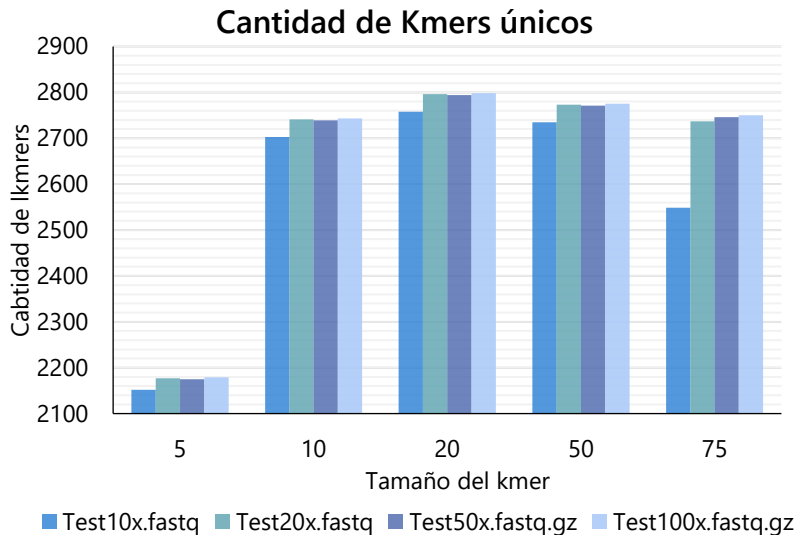
	Secuencia	Tiempo (ms)	Rango overl.
Test10x.fastq	I peloton de fusilamiento el coronel Aureliano Buendia habia de recordar aquella tarde remota en que su padre lo llevo a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y canhabrava construidas a la orilla de un rio de aguas diafanas que se precipitaban por un lecho de piedras pulidas blancas y enormes como huevos prehistoricos. El mundo era tan reciente que muchas cosas carecian de nombre y para mencionarlas habia que senhalarlas con el dedo. Todos los anhos por el mes de marzo una familia de gitanos desarrapados plantaba su carpa cerca de la aldea y con un grande alboroto de pitos y timbales daban a conocer los nuevos inventos. Primero llevaron el iman. Un gitano corpulento de barba montaraz y manos de gorrión que se presento con el nombre de Melquiades hizo una truculenta demostracion publica de lo que el mismo llamaba la octava maravilla de los sabios alquimistas de Macedonia. Fue de casa en casa arrastrando dos lingotes metalicos y todo el mundo se espanto al ver que los calderos las pailas las tenazas y los anafes se caian de su sitio y las maderas crujian por la desesperacion de los clavos y los tornillos tratando de desenclavarse y aun los objetos perdidos desde hacia mucho tiempo aparecian por donde mas se les habia buscado y se arrastraban en desbandada turbulenta detras de los fierros magicos de Melquiades. Las cosas tienen vida propia - pregonaba el gitano con aspero acento - todo es cuestion de despertarles el anima. Jose Arcadio Buendia cuya desaforada imaginacion iba siempre mas lejos que el ingenio de la naturaleza y aun mas alla del milagro y la magia penso que era posible servirse de aquella invencion inutil para desentranhar el oro de la tierra. Melquiades - que era un hombre honrado - le previno: Para eso no sirve. Pero Jose Arcadio Buendia no creia en aquel tiempo en la honradez de los gitanos así que cambio su mulo y una partida de chivos por los dos lingotes imantados. Ursula Iguaran su mujer que contaba con aquellos animales para ensanchar el desmedrado patrimonio domestico no consiguio disuadirlo. Muy pronto ha de sobrnarnos oro para empedrar la casa replico su marido. Durante varios meses se empenho en demostrar el acierto de sus conjeturas. Exploro palmo a palmo la region inclusive el fondo del rio arrastrando los dos lingotes de hierro y recitando en voz alta el conjuro de Melquiades. Lo unico que logro desenterrar fue una armadura del siglo XV con todas sus partes soldadas por un cascote de oxido cuyo interior tenia la resonancia hueca de un enorme calabazo lleno de piedras. Cuando Jose Arcadio Buendia y los cuatro hombres de su expedicion lograron desarticular la armadura encontraron dentro un esqueleto calcificado que llevaba colgado en el cuello un relicario de cobre con un rizo	14	2-50
Test20x.fastq	chos anhos despues frente al peloton de fusilamiento el coronel Aureliano Buendia habia de recordar aquella tarde remota en que su padre lo llevo a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y canhabrava construidas a la orilla de un rio de aguas diafanas que se precipitaban por un lecho de piedras pulidas blancas y enormes como huevos prehistoricos. El mundo era tan reciente que muchas cosas carecian de nombre y para mencionarlas habia que senhalarlas con el dedo. Todos los anhos por el mes de marzo una familia de gitanos desarrapados plantaba su carpa cerca de la aldea y con un grande alboroto de pitos y timbales daban a conocer los nuevos inventos. Primero llevaron el iman. Un gitano corpulento de barba montaraz y manos de gorrión que se presento con el nombre de Melquiades hizo una truculenta demostracion publica de lo que el mismo llamaba la octava maravilla de los sabios alquimistas de Macedonia. Fue de casa en casa arrastrando dos lingotes metalicos y todo el mundo se espanto al ver que los calderos las pailas las tenazas y los anafes se caian de su sitio y las maderas crujian por la desesperacion de los clavos y los tornillos tratando de desenclavarse y aun los objetos perdidos desde hacia mucho tiempo aparecian por donde mas se les habia buscado y se arrastraban en desbandada turbulenta detras de los fierros magicos de Melquiades. Las cosas tienen vida propia - pregonaba el gitano con aspero acento - todo es cuestion de despertarles el anima. Jose Arcadio Buendia cuya desaforada imaginacion iba siempre mas lejos que el ingenio de la naturaleza y aun mas alla del milagro y la magia penso que era posible servirse de aquella invencion inutil para desentranhar el oro de la tierra. Melquiades - que era un hombre honrado - le previno: Para eso no sirve. Pero Jose Arcadio Buendia no creia en aquel tiempo en la honradez de los gitanos así que cambio su mulo y una partida de chivos por los dos lingotes imantados. Ursula Iguaran su mujer que contaba con aquellos animales para ensanchar el desmedrado patrimonio domestico no consiguio disuadirlo. Muy pronto ha de sobrnarnos oro para empedrar la casa replico su marido. Durante varios meses se empenho en demostrar el acierto de sus conjeturas. Exploro palmo a palmo la region inclusive el fondo del rio arrastrando los dos lingotes de hierro y recitando en voz alta el conjuro de Melquiades. Lo unico que logro desenterrar fue una armadura del siglo XV con todas sus partes soldadas por un cascote de oxido cuyo interior tenia la resonancia hueca de un enorme calabazo lleno de piedras. Cuando Jose Arcadio Buendia y los cuatro hombres de su expedicion lograron desarticular la armadura encontraron dentro un esqueleto calcificado que llevaba colgado en el cuello un relicario de cobre con un rizo de mujer.	27	2- 70

Test50x.fastq	<p>os años después frente al pelotón de fusilamiento el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y canchales construidas a la orilla de un río de aguas diáfanas que se precipitaban por un lecho de piedras pulidas, blancas y enormes como huevos prehistóricos. El mundo era tan reciente que muchas cosas carecían de nombre y para mencionarlas había que señalarlas con el dedo. Todos los años por el mes de marzo una familia de gitanos desarrapados plantaba su carpa cerca de la aldea y con un grande alboroto de pitos y timbales daban a conocer los nuevos inventos. Primero llevaron el imán. Un gitano corpulento de barba montaraz y manos de gorrión que se presentó con el nombre de Melquíades hizo una truculenta demostración pública de lo que él mismo llamaba la octava maravilla de los sabios alquimistas de Macedonia. Fue de casa en casa arrastrando dos lingotes metálicos y todo el mundo se espantó al ver que los calderos, las pailas, las tenazas y los ánafes se caían de su sitio y las maderas crujían por la desesperación de los clavos y los tornillos tratando de desenclavar y aun los objetos perdidos desde hacía mucho tiempo aparecían por donde más se les había buscado y se arrastraban en desbandada turbulenta detrás de los fierros mágicos de Melquíades. Las cosas tienen vida propia - pregonaba el gitano con áspero acento - todo es cuestión de despertarles el alma. José Arcadio Buendía cuya desahogada imaginación iba siempre más lejos que el ingenio de la naturaleza y aun más allá del milagro y la magia pensó que era posible servirse de aquella invención inútil para desentrañar el oro de la tierra. Melquíades - que era un hombre honrado - le previno: Para eso no sirve. Pero José Arcadio Buendía no creía en aquel tiempo en la honradez de los gitanos así que cambió su mulo y una partida de chivos por los dos lingotes imantados. Úrsula Iguarán su mujer que contaba con aquellos animales para ensanchar el desmedrado patrimonio doméstico no consiguió disuadirlo. Muy pronto ha de sobrarnos oro para empedrar la casa, replicó su marido. Durante varios meses se empeñó en demostrar el acierto de sus conjeturas. Exploró palmo a palmo la región inclusive el fondo del río arrastrando los dos lingotes de hierro y recitando en voz alta el conjuro de Melquíades. Lo único que logró desenterrar fue una armadura del siglo XV con todas sus partes soldadas por un cascote de óxido cuyo interior tenía la resonancia hueca de un enorme calabazo lleno de piedras. Cuando José Arcadio Buendía y los cuatro hombres de su expedición lograron desarticular la armadura encontraron dentro un esqueleto calcificado que llevaba colgado en el cuello un relicario de cobre con un rizo de mujer.</p>	49	10-87
Test100x.fastq	<p>Muchos años después frente al pelotón de fusilamiento el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y canchales construidas a la orilla de un río de aguas diáfanas que se precipitaban por un lecho de piedras pulidas, blancas y enormes como huevos prehistóricos. El mundo era tan reciente que muchas cosas carecían de nombre y para mencionarlas había que señalarlas con el dedo. Todos los años por el mes de marzo una familia de gitanos desarrapados plantaba su carpa cerca de la aldea y con un grande alboroto de pitos y timbales daban a conocer los nuevos inventos. Primero llevaron el imán. Un gitano corpulento de barba montaraz y manos de gorrión que se presentó con el nombre de Melquíades hizo una truculenta demostración pública de lo que él mismo llamaba la octava maravilla de los sabios alquimistas de Macedonia. Fue de casa en casa arrastrando dos lingotes metálicos y todo el mundo se espantó al ver que los calderos, las pailas, las tenazas y los ánafes se caían de su sitio y las maderas crujían por la desesperación de los clavos y los tornillos tratando de desenclavar y aun los objetos perdidos desde hacía mucho tiempo aparecían por donde más se les había buscado y se arrastraban en desbandada turbulenta detrás de los fierros mágicos de Melquíades. Las cosas tienen vida propia - pregonaba el gitano con áspero acento - todo es cuestión de despertarles el alma. José Arcadio Buendía cuya desahogada imaginación iba siempre más lejos que el ingenio de la naturaleza y aun más allá del milagro y la magia pensó que era posible servirse de aquella invención inútil para desentrañar el oro de la tierra. Melquíades - que era un hombre honrado - le previno: Para eso no sirve. Pero José Arcadio Buendía no creía en aquel tiempo en la honradez de los gitanos así que cambió su mulo y una partida de chivos por los dos lingotes imantados. Úrsula Iguarán su mujer que contaba con aquellos animales para ensanchar el desmedrado patrimonio doméstico no consiguió disuadirlo. Muy pronto ha de sobrarnos oro para empedrar la casa, replicó su marido. Durante varios meses se empeñó en demostrar el acierto de sus conjeturas. Exploró palmo a palmo la región inclusive el fondo del río arrastrando los dos lingotes de hierro y recitando en voz alta el conjuro de Melquíades. Lo único que logró desenterrar fue una armadura del siglo XV con todas sus partes soldadas por un cascote de óxido cuyo interior tenía la resonancia hueca de un enorme calabazo lleno de piedras. Cuando José Arcadio Buendía y los cuatro hombres de su expedición lograron desarticular la armadura encontraron dentro un esqueleto calcificado que llevaba colgado en el cuello un relicario de cobre con un rizo de mujer.</p>	69	2-92

El algoritmo se demora más tiempo ensamblando la secuencia que tiene una mayor profundidad. De igual forma, el rango de solapamientos que necesita cada uno de los archivos para poder realizar el ensamblaje de manera correcta aumenta a medida que aumenta la profundidad de las secuencias.

Resultados obtenidos de Kmers

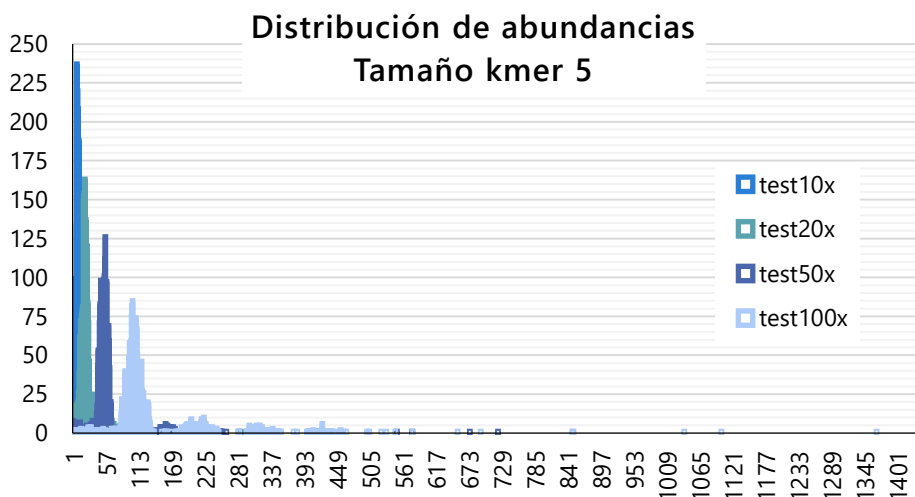
Cantidad de Kmers únicos



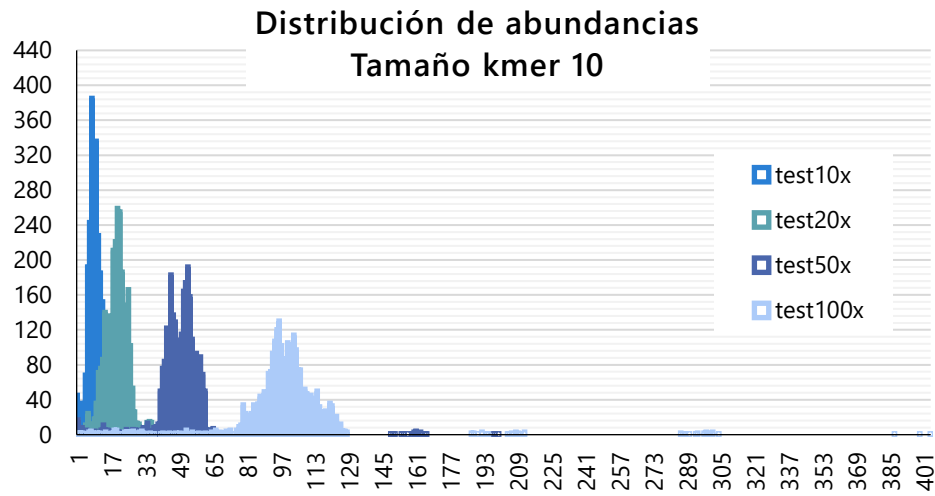
Gráfica 3 Cantidad de Kmers únicos. Los datos con los que se hizo la gráfica se encuentran en la Tabla 6

A medida que aumenta el tamaño del kmer, disminuye la cantidad de kmers **totales**. Sin embargo, la cantidad de kmers **únicos** depende de igual forma de la probabilidad de que hayan o no kmers **iguales**, por ejemplo, es mucho más probable encontrar kmers con tamaño 5bp que sean iguales (hay $5^4 = 625$ posibilidades), mientras que cuando los kmers tienen otro tamaño, es mucho menos probable ($10^4 = 10000$, $20^4 = 160000$, $50^4 = 6250000$, $75^4 = 31640625$). Debido a esto, en todos los archivos es posible encontrar que hayan mas kmers únicos con un tamaño de 20bp

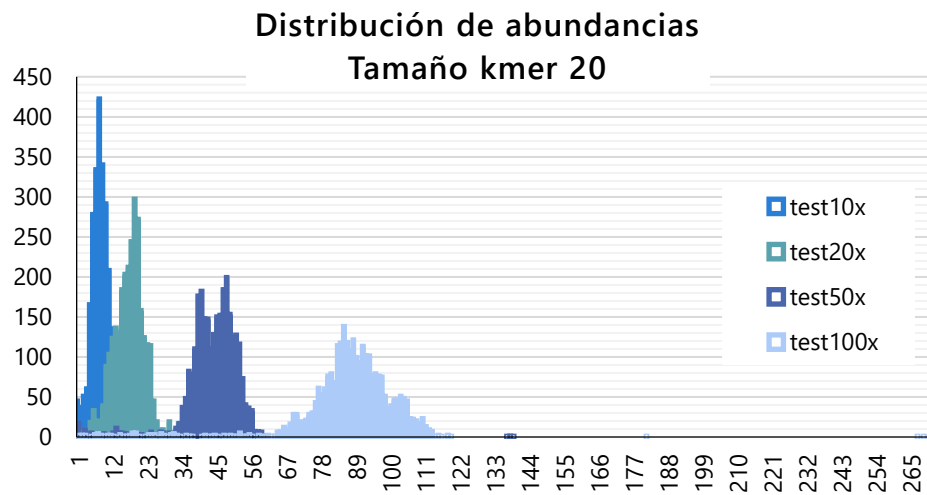
Abundancia media



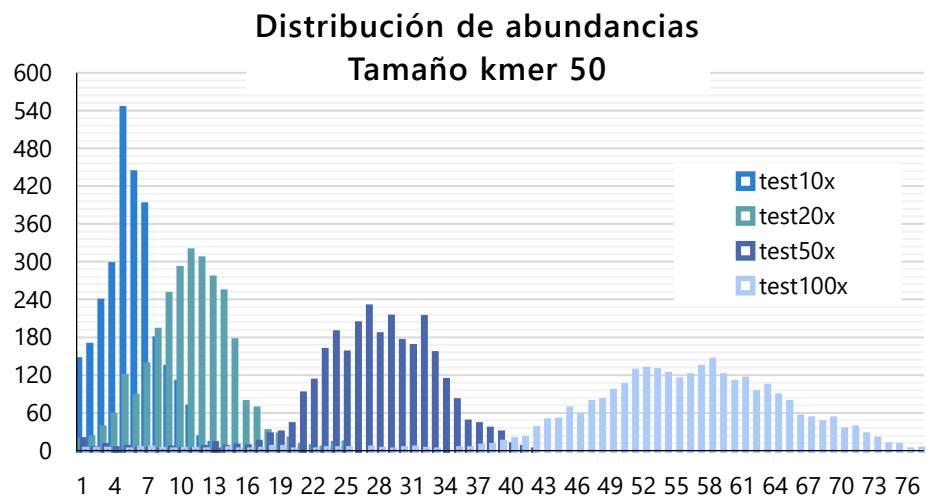
Gráfica 4 Distribución de abundancias con un tamaño de kmer de 5



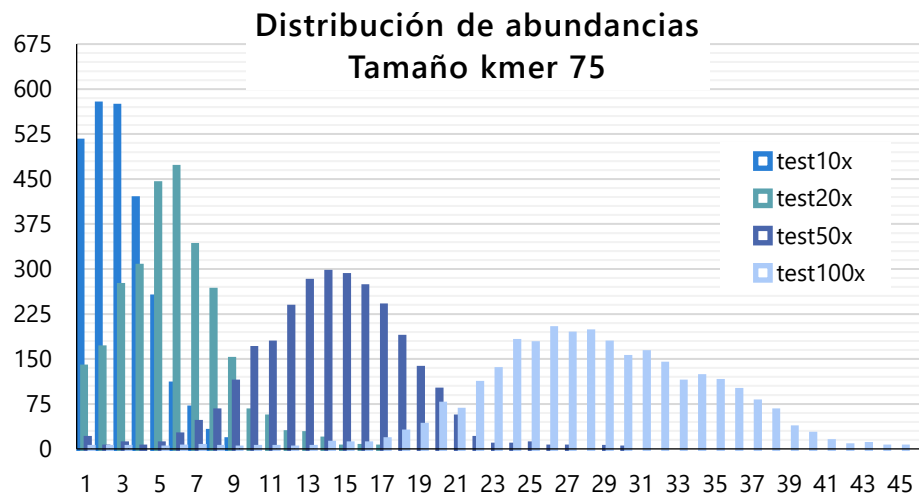
Gráfica 5 Distribución de abundancias con un tamaño de kmer de 10



Gráfica 6 Distribución de abundancias con un tamaño de kmer de 20

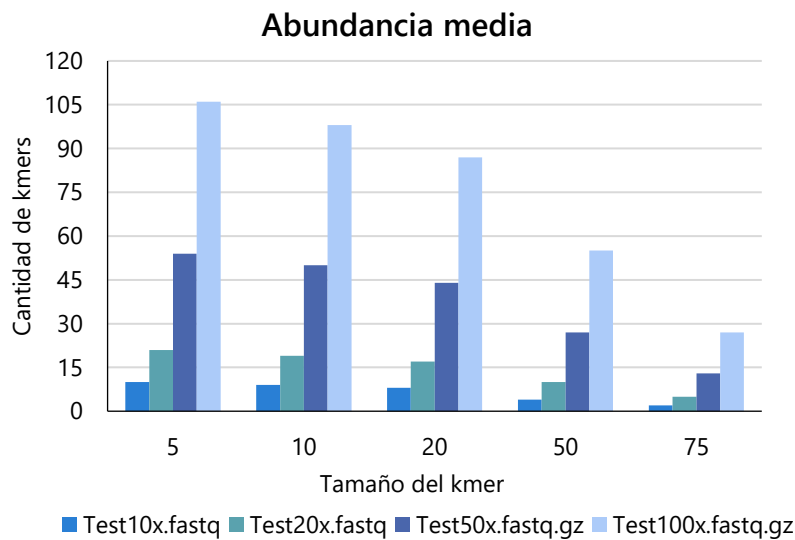


Gráfica 7 Distribución de abundancias con un tamaño de kmer de 50



Gráfica 8 Distribución de abundancias con un tamaño de kmer de 75

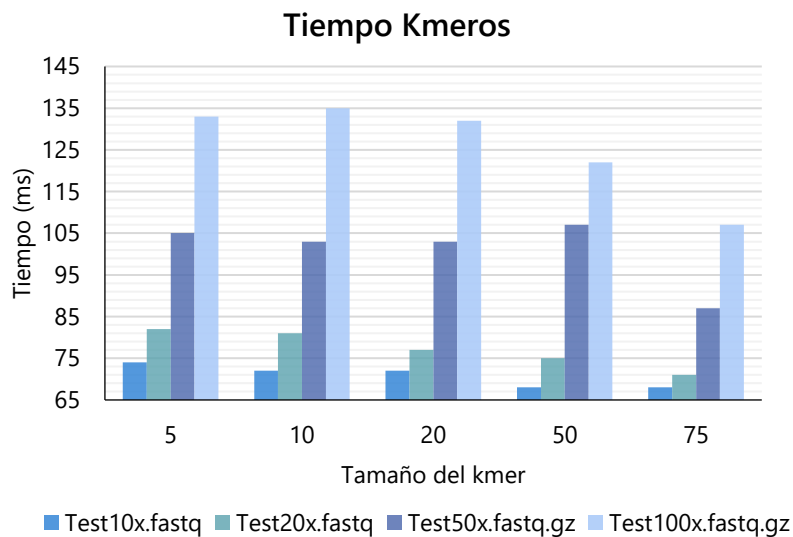
De acuerdo con la Gráfica 4, es importante observar que cuando el kmer es igual a 5, la cantidad de kmers reeptidos es superior a cuando los kmers toman otro tamaño. Esto se debe a que, como ya se mencionó anteriormente, es mucho más probable encontrar kmers con tamaño 5bp que sean iguales, a diferencia de las otras gráficas en donde la distribución de abundancias tiene un rango menor.



Gráfica 9 Abundancia media de los kmeros. Los datos para hacer la gráfica se encuentran en la Tabla 7

En la Gráfica 9, es posible observar que la abundancia media aumenta a medida que aumenta la profundidad y disminuye a medida que aumenta el tamaño del kmer.

Tiempo que se demora en construir la tabla con los kmeros



Gráfica 10 Tiempo que se demora en construir la tabla de los kmeros. Los datos para hacer esta gráfica se encuentran en la Tabla 8.

De acuerdo con la gráfica anterior, se demora menos tiempo construyendo los kmeros de tamaño 75pb, sin embargo, la diferencia se nota mas en los archivos con profundidad mayor. Es decir que el tiempo aumenta cuando el tamaño de los kmers disminuye.

Resultados de SimpleReadsSimulator

Para realizar la simulación de la lectura de secuencias aleatorias a partir de una secuencia, se utilizó la secuencia **Zebrafish DNA sequence from clone DKEY-208G1 in linkage group 5, complete sequence** de NCBI (<https://www.ncbi.nlm.nih.gov/nucore/BX537263>). La secuencia tiene una longitud de 166940 bp, y se eliminaron algunas bases para que quedara de 20000 bp (data/sequence3.fasta) por lo que para la construcción de lecturas aleatorias se utilizó la siguiente cantidad de lecturas:

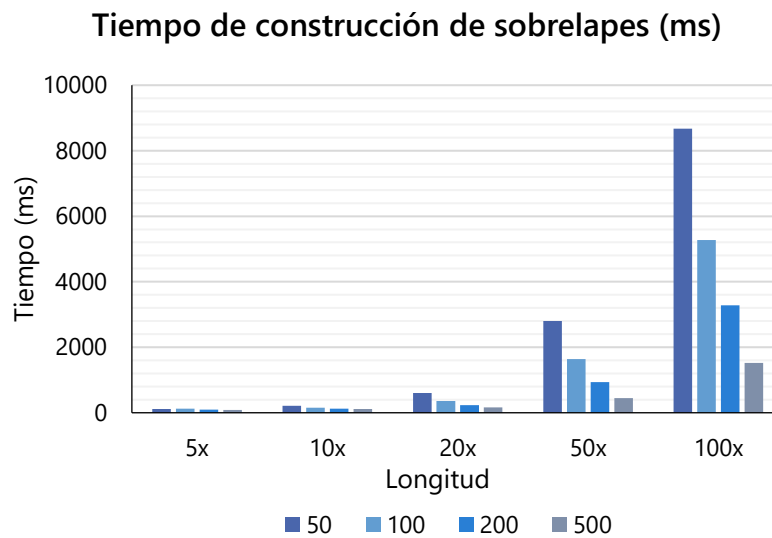
Longitud lectura	Profundidad	Cantidad de lecturas
50	5	2000
100	5	1000
200	5	500
500	5	200
50	10	4000
100	10	2000
200	10	1000
500	10	400
50	20	8000
100	20	4000
200	20	2000
500	20	800
50	50	20000
100	50	10000

200	50	5000
500	50	2000
50	100	40000
100	100	20000
200	100	10000
500	100	4000

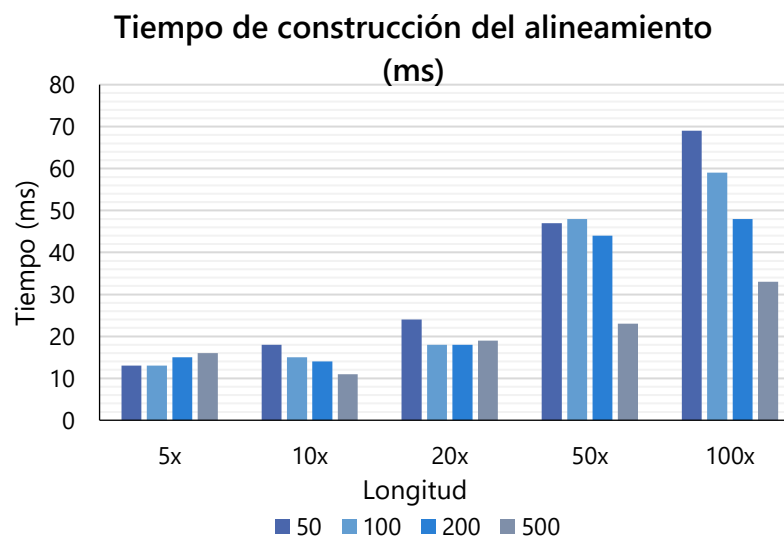
La cantidad de lecturas se calculo con la siguiente formula:

$$\text{numReads} = \frac{\text{Profundad} \times \text{seqLength}}{\text{readLength}}$$

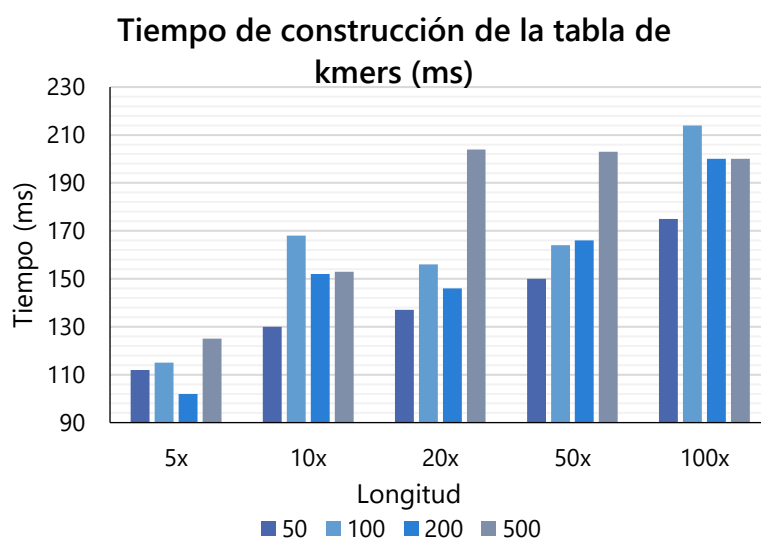
Los resultados de los tiempos se muestran a continuación:



Gráfica 11 Tiempo (ms) que se demora en la construcción de la tabla de sobrelapes. Los datos para hacer la gráfica se encuentran en la Tabla 9



Gráfica 12 Tiempo (ms) que se demora en la construcción del alineamiento. Los datos para hacer la gráfica se encuentran en la Tabla 10

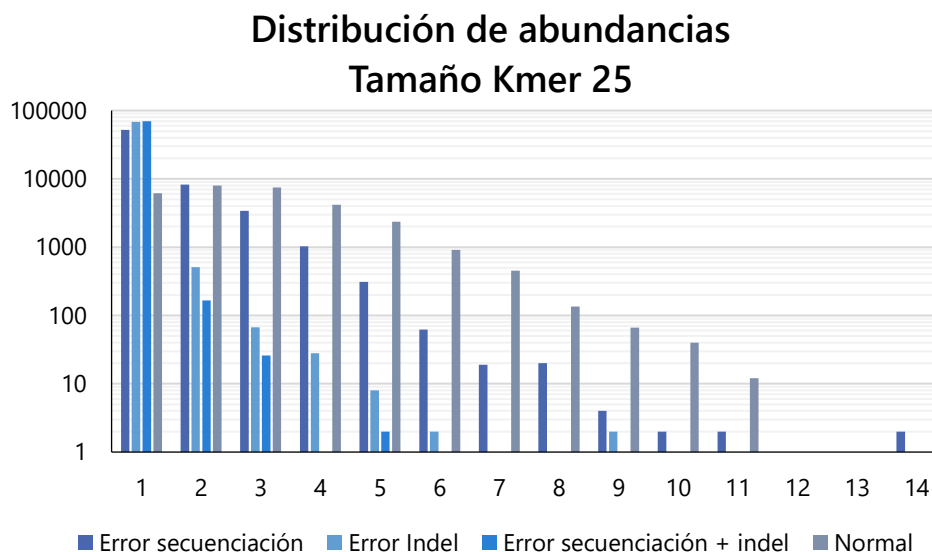


Gráfica 13 Tiempo (ms) que se demora en la construcción de la tabla de los kmers. Los datos para hacer la gráfica se encuentran en la Tabla 11

En las Gráfica 11- Gráfica 12 se puede observar que los tiempos que se gastan corriendo los dos algoritmos de OverlapGraph incrementa cuando se incrementa la profundidad de las lecturas, pero disminuyen con la longitud de la lectura. En la Gráfica 13 se observa un comportamiento similar en lo que respecta a la profundidad.

Generación de errores

Para la generación de errores se utilizaron dos argumentos: La tasa de errores de secuenciación (como 0.02) y la tasa de indels (0.1), los resultados se muestran a continuación:



Gráfica 14: Distribución de abundancias. Los datos para hacer la gráfica se

Cuando existen errores, hay mas kmers únicos en las secuencias que tienen los errores inducidos. Sin embargo, debido a que esta secuencia se genero con valores de una secuencia real, es posible que esta también maneje algunos errores.

Tablas adjuntas

Tabla 3 Resultados de la distribución de abundancias de secuencias

	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
1	270	487	874	1012
2	15	52	233	546
3		3	44	197
4			7	62
5				9
6				2

Tabla 4 Distribución de sucesores por secuencia

	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
0	0	0	0	0
1	0	0	0	0
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	2	1	1	1
6	4	1	1	1
7	15	1	1	1
8	24	2	1	1
9	31	3	1	1
10	32	4	1	1
11	37	5	1	1
12	38	12	1	1
13	27	17	1	1
14	20	19	1	1
15	16	26	1	1
16	13	31	1	1
17	15	33	1	1
18	5	47	1	1
19	2	57	1	1
20	1	49	1	1
21		49	1	1
22		50	1	1
23		37	1	1
24		31	1	1
25		30	1	1
26		10	1	1
27		9	1	1
28		5	1	1
29		2	1	1
30		3	7	1
31		5	4	1
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				

66			111
67			84
68			62
69			38
70			75
71			47
72			29
73			3
74			4

75			1
76			3
77			0
78			1
79			0
80			0
81			0
82			1

Tabla 5 Cantidad de kmers totales

Tamaño de Kmer	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
5	28800	57600	144000	288000
10	27300	54600	136500	273000
20	24300	48600	121500	243000
50	15300	30600	76500	153000
75	7800	15600	39000	78000

Tabla 6 Cantidad de kmers Unicos

Tamaño de Kmer	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
5	2152	2177	2175	2179
10	2703	2741	2739	2743
20	2758	2796	2794	2798
50	2735	2773	2771	2775
75	2549	2737	2746	2750

Tabla 7 Abundancia media de Kmeros

Tamaño de Kmer	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
5	10	21	54	106
10	9	19	50	98
20	8	17	44	87
50	4	10	27	55
75	2	5	13	27

Tabla 8 Tiempo que demora en construir la Tabla de kmeros

Tamaño de Kmer	Test10x.fastq	Test20x.fastq	Test50x.fastq.gz	Test100x.fastq.gz
5	74	82	105	133
10	72	81	103	135
20	72	77	103	132
50	68	75	107	122
75	68	71	87	107

Tabla 9 Tiempo (ms) que se demora construyendo la tabla de los sobrelapes

Tamaño /Profundidad	50	100	200	500
---------------------	----	-----	-----	-----

5x	119	120	92	89
10x	216	150	127	115
20x	606	360	235	160
50x	2804	1640	940	452
100x	8668	5274	3281	1522

Tabla 10 Tiempo (ms) que se demora haciendo el ensamblaje de la secuencia

Tamaño /Profundidad	50	100	200	500
5x	-	13	15	-
10x	18	15	14	11
20x	24	18	18	19
50x	47	48	44	23
100x	69	59	48	33

Tabla 11 Tiempo (ms) que se demora en la construcción de kmers

Tamaño /Profundidad	50	100	200	500
5x	112	115	102	125
10x	130	168	152	153
20x	137	156	146	204
50x	150	164	166	203
100x	175	214	200	200

Tabla 12 Resultado distribución kmers errores secuenciación

	Error secuenciación	Error Indel	Error secuenciación + indel	Normal
1	52306	68576	69573	6173
2	8256	508	166	8008
3	3389	67	26	7456
4	1022	28	0	4163
5	309	8	2	2369
6	62	2		908
7	19	0		451
8	20	1		135
9	4	2		66
10	2	1		40
11	2			12
12	0			1
13	0			1
14	2			1
15	1			0
16	1			2
17	0			0
18	0			0
19	1			0
20	0			0
21	0			0

22	1			0
23	1			0
24	0			0
25	0			0
26	0			1
27	0			2
28	0			1
29	0			0
30	0			1
31	0			0
32	0			0
33	0			1
34	0			1
35	0			0
36	0			0
37	0			0
38	0			0
39	0			0
40	1			0
41	0			0
42	0			0
43	2			0
44	1			0
45	0			0
46	0			0
47	0			0
48	0			0
49	0			0
50	0			0
51	0			0
52	0			0
53	0			0
54	0			0
55	0			0
56	0			0
57	0			0
58	0			1