

Tarea 2 - Ensamblaje de Genomas

8 de septiembre de 2018

Stephannie Jiménez
Michael Medina

Código: 201423727
Código: 201727325

1. [30 %] Grafo de sobrelapes

A continuación se muestran la distribución de abundancias de secuencias, la distribución de sucesores por secuencia y el tiempo de ejecución para cada uno de los ejemplos dados.

1.1. test10x

En este archivo, se distinguieron 146 secuencias diferentes y el algoritmo se demoró 45ms en construir el grafo. Las figuras 1 y 2 muestran las distribuciones obtenidas para la abundancia de secuencias y la de sucesores por secuencia.

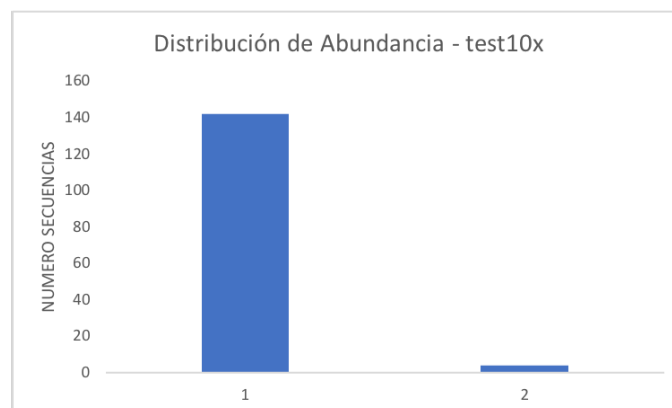


Figura 1: Distribución de abundancias de secuencias para la muestra *test10x*.

1.2. test20x

Para esta entrada, se encontraron 277 secuencias diferentes y se logró armar el grafo de sobrelapes en 64 ms. Las figuras 3 y 4 muestran las distribuciones obtenidas para la abundancia de secuencias y la de sucesores por secuencia.

1.3. test50x

De igual forma, se determinó que en este archivo existen 597 secuencias diferentes y se construyó el grafo de sobrelapes en 150ms. Las figuras 5 y 6 muestran las distribuciones obtenidas para la abundancia de secuencias y la de sucesores por secuencia.

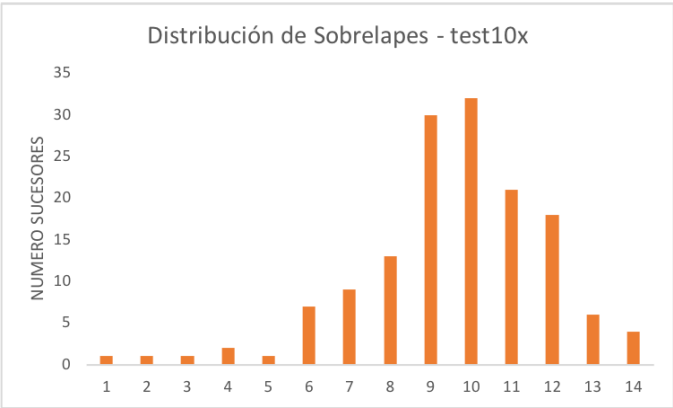


Figura 2: Distribución de sucesores por secuencia para la muestra *test10x*.

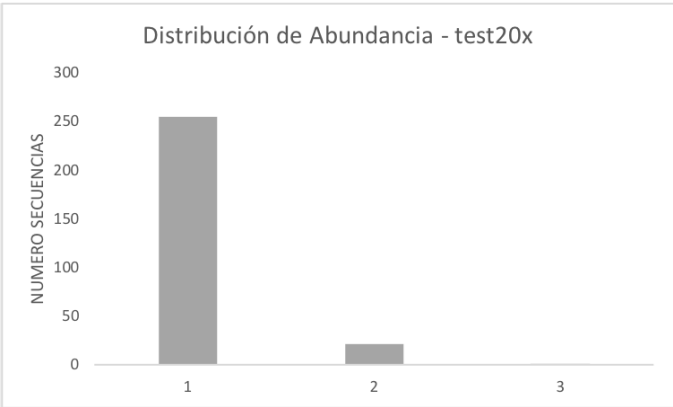


Figura 3: Distribución de abundancias de secuencias para la muestra *test20x*.

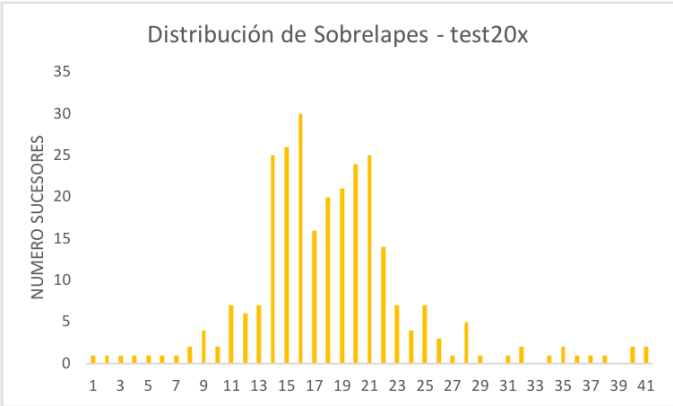


Figura 4: Distribución de sucesores por secuencia para la muestra *test20x*.

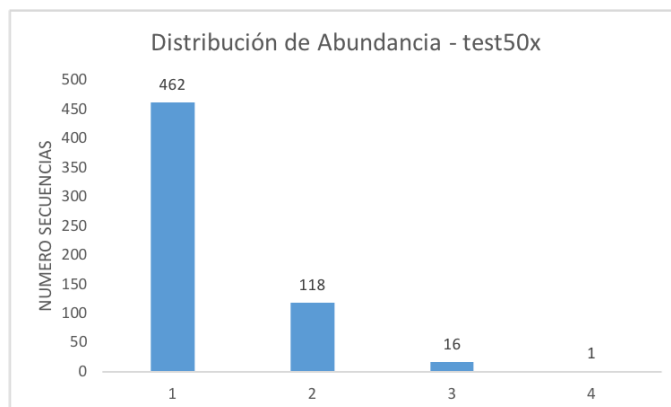


Figura 5: Distribución de abundancias de secuencias para la muestra *test50x*.

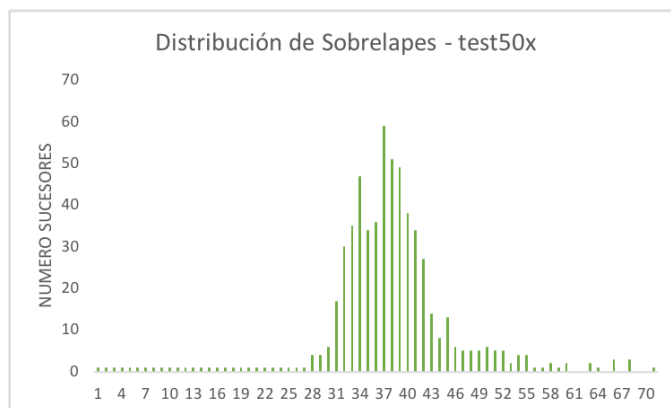


Figura 6: Distribución de sucesores por secuencia para la muestra *test50x*.

1.4. test100x

Por último, se utilizó este archivo donde se encontraron 928 secuencias diferentes y el grafo se armó en 264ms. Las figuras 7 y 8 muestran las distribuciones obtenidas para la abundancia de secuencias y la de sucesores por secuencia.

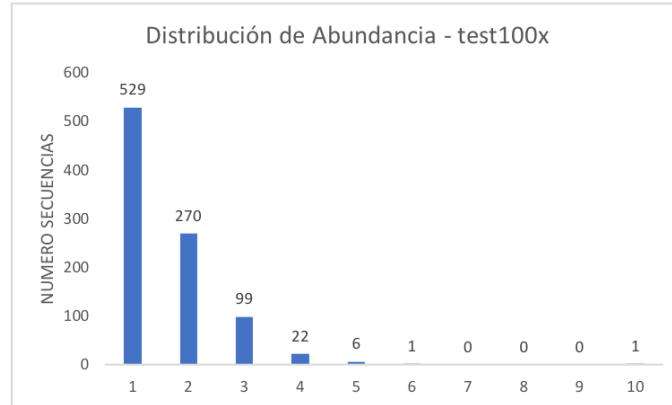


Figura 7: Distribución de abundancias de secuencias para la muestra *test100x*.

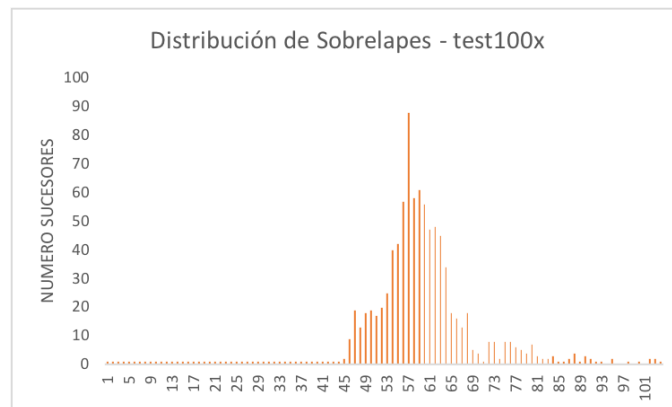


Figura 8: Distribución de sucesores por secuencia para la muestra *test100x*.

2. [30 %] K-mers

Se probó con cada muestra diferentes tamaños de k-mer registrando la cantidad de k-mers, la abundancia media y el tiempo de ejecución. Enseguida se muestran los resultados obtenidos. Para cada uno de los archivos de prueba, se varió el parámetro del tamaño de k-mer con valores de 5, 10, 20, 50 y 75.

2.1. test10x

En las figuras 9 a 13 y la tabla 1 se encuentran los resultados para este archivo.

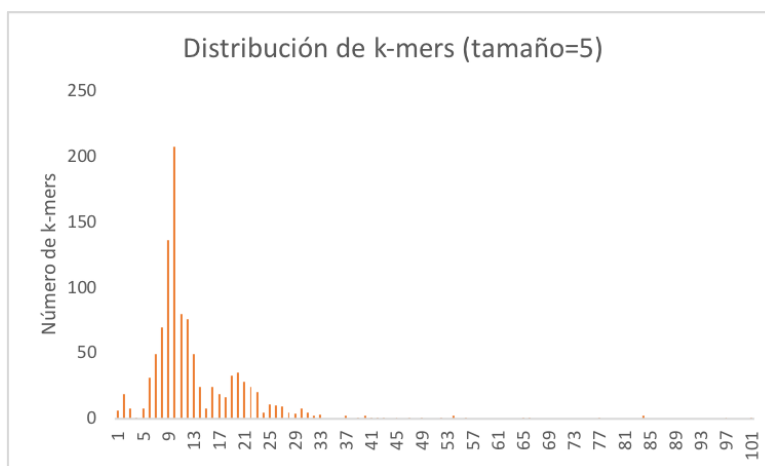


Figura 9: Distribución de k-mers de tamaño 5 para la muestra *test10x*.

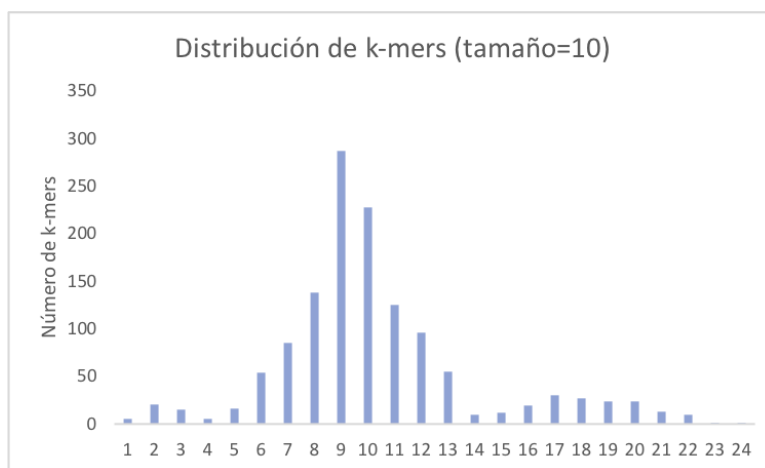


Figura 10: Distribución de k-mers de tamaño 10 para la muestra *test10x*.

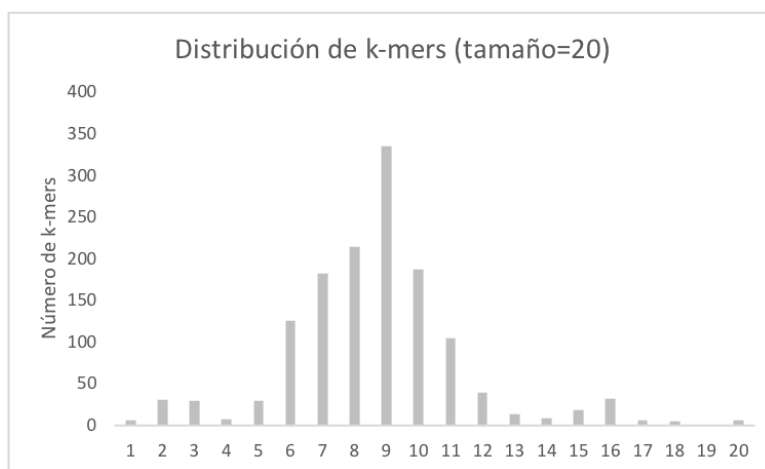


Figura 11: Distribución de k-mers de tamaño 20 para la muestra *test10x*.

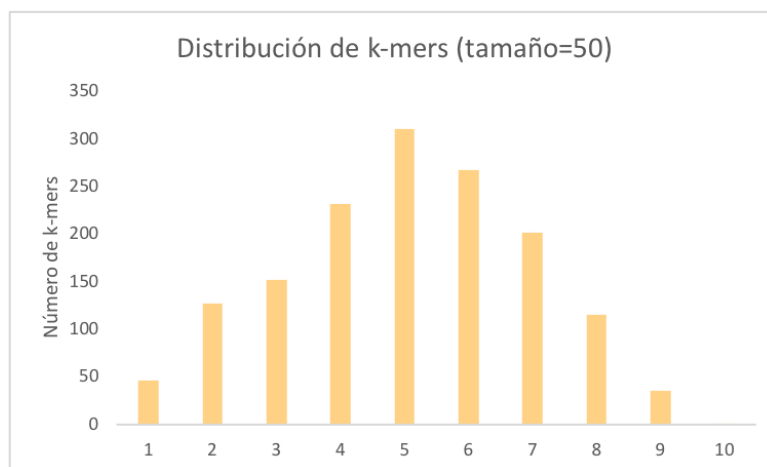


Figura 12: Distribución de k-mers de tamaño 50 para la muestra *test10x*.

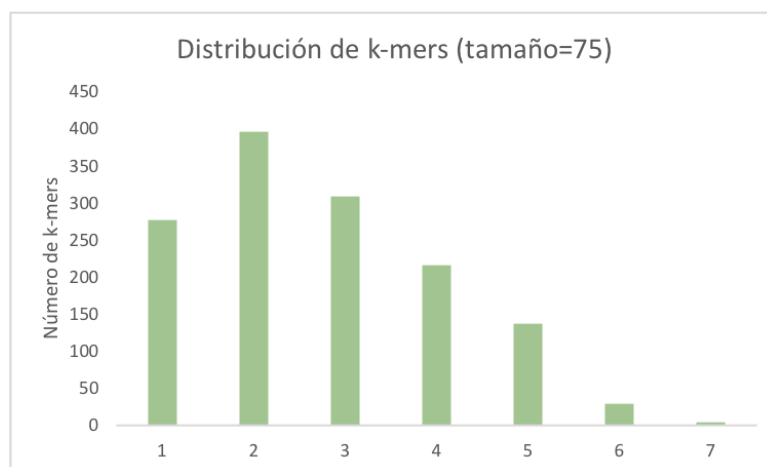


Figura 13: Distribución de k-mers de tamaño 75 para la muestra *test10x*.

Cuadro 1: Resultados de las pruebas realizadas variando el tamaño de los k-mers generados.

	5	10	20	50	75
Número de k-mers totales	1056	1311	1384	1488	1370
Tiempo (ms)	38	39	40	34	33
Abundancia media	10	9	9	5	2

2.2. test20x

En las figuras 14 a 15 y la tabla 2 se encuentran los resultados para este archivo.

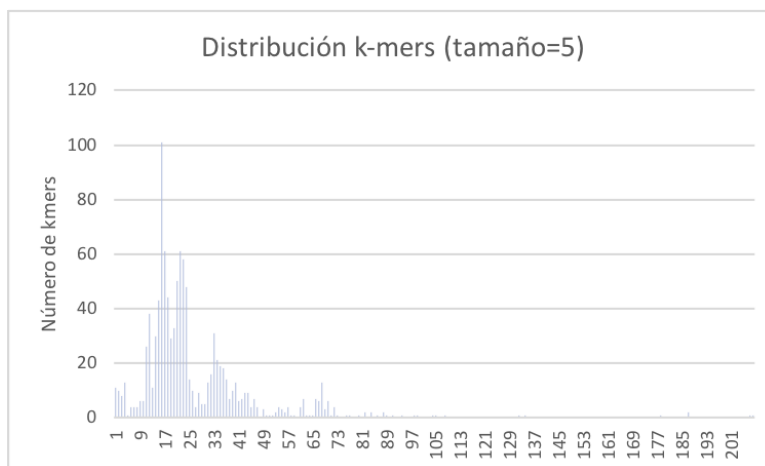


Figura 14: Distribución de k-mers de tamaño 5 para la muestra *test20x*.

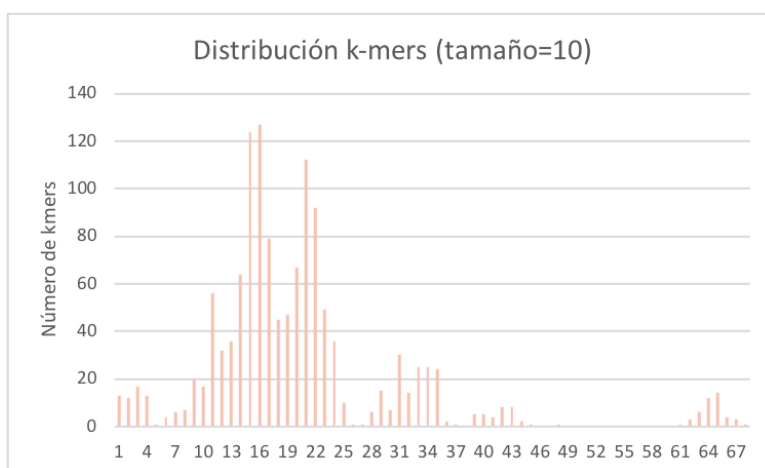


Figura 15: Distribución de k-mers de tamaño 10 para la muestra *test20x*.

Cuadro 2: Resultados de las pruebas realizadas variando el tamaño de los k-mers generados.

	5	10	20	50	75
Número de k-mers totales	1058	1315	1388	1495	1477
Tiempo (ms)	58	50	50	44	42
Abundancia media	16	16	13	11	4

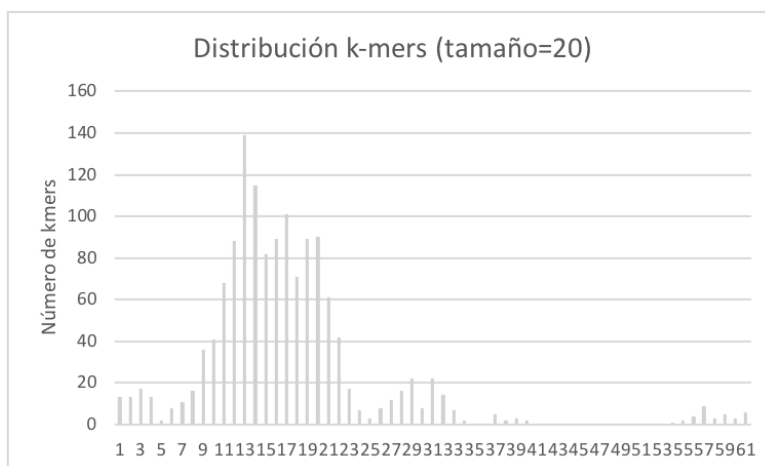


Figura 16: Distribución de k-mers de tamaño 20 para la muestra *test20x*.

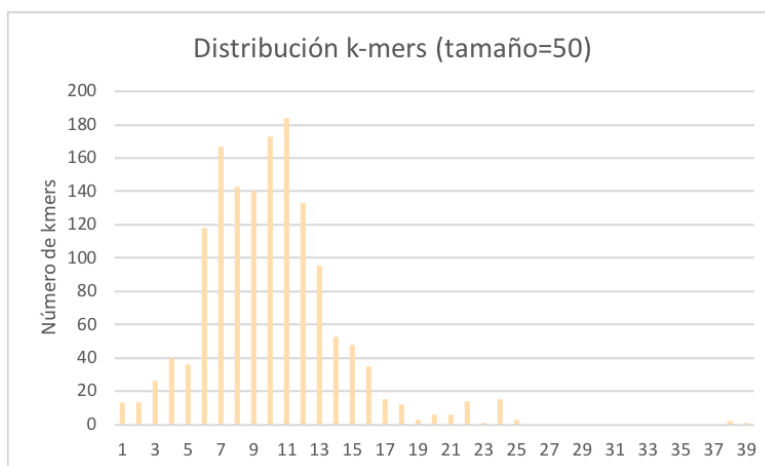


Figura 17: Distribución de k-mers de tamaño 50 para la muestra *test20x*.

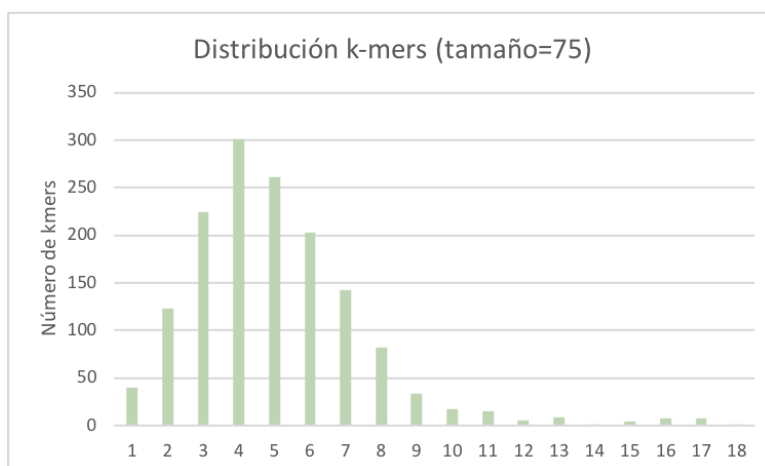


Figura 18: Distribución de k-mers de tamaño 75 para la muestra *test20x*.

2.3. test50x

En las figuras 19 a 23 y la tabla 3 se encuentran los resultados para este archivo.

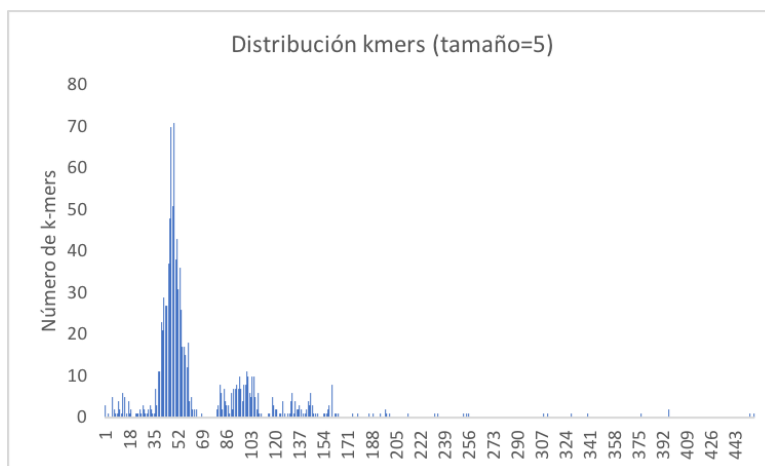


Figura 19: Distribución de k-mers de tamaño 5 para la muestra *test50x*.

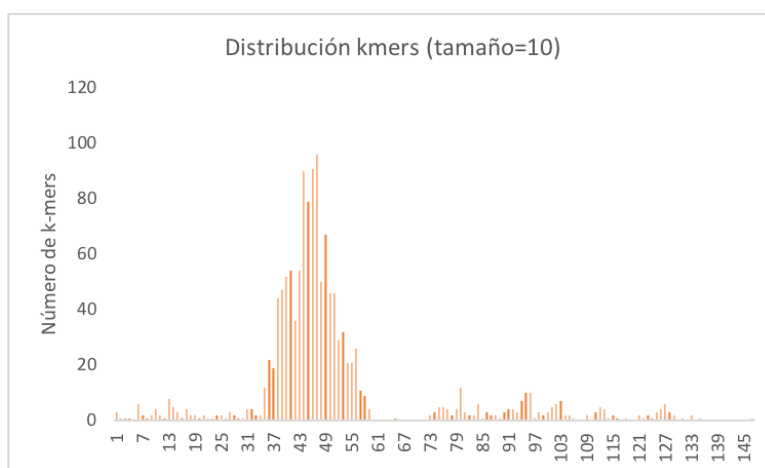


Figura 20: Distribución de k-mers de tamaño 10 para la muestra *test50x*.

Cuadro 3: Resultados de las pruebas realizadas variando el tamaño de los k-mers generados.

	5	10	20	50	75
Número de k-mers totales	1059	1317	1390	1497	1479
Tiempo (ms)	67	88	83	97	67
Abundancia media	49	47	41	24	11

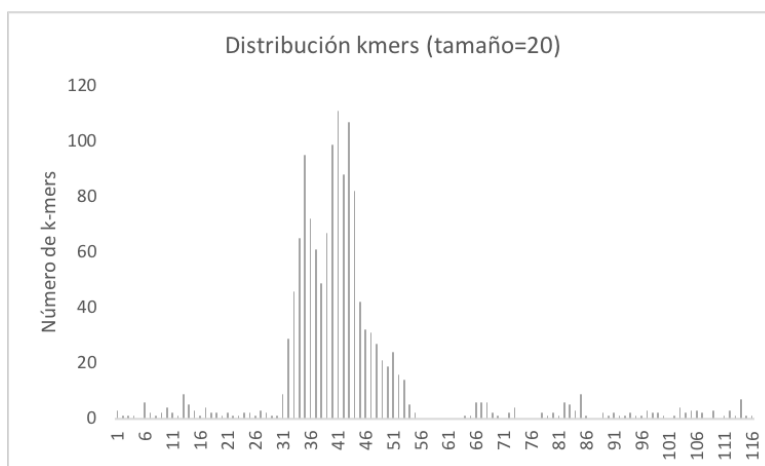


Figura 21: Distribución de k-mers de tamaño 20 para la muestra *test50x*.

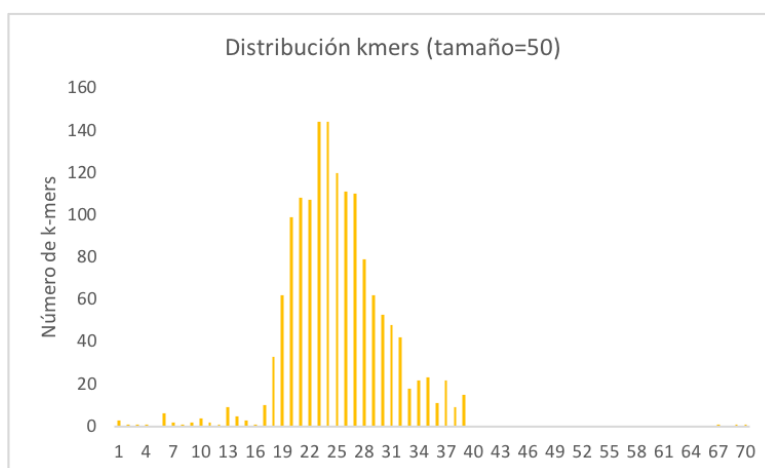


Figura 22: Distribución de k-mers de tamaño 50 para la muestra *test50x*.

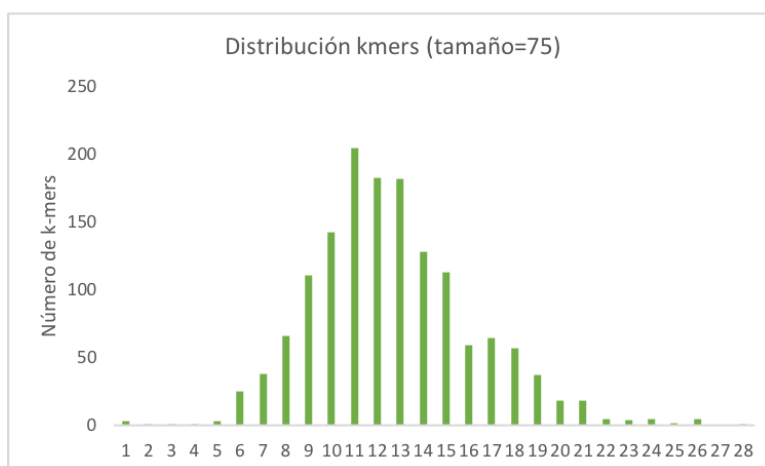


Figura 23: Distribución de k-mers de tamaño 75 para la muestra *test50x*.

2.4. test100x

En las figuras 24 a 28 y la tabla 4 se encuentran los resultados para este archivo.

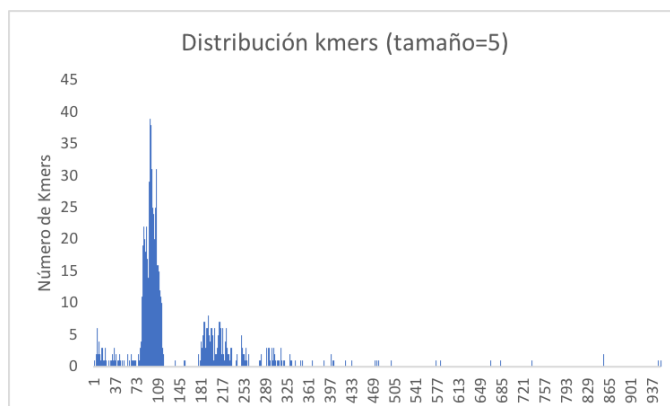


Figura 24: Distribución de k-mers de tamaño 5 para la muestra *test100x*.

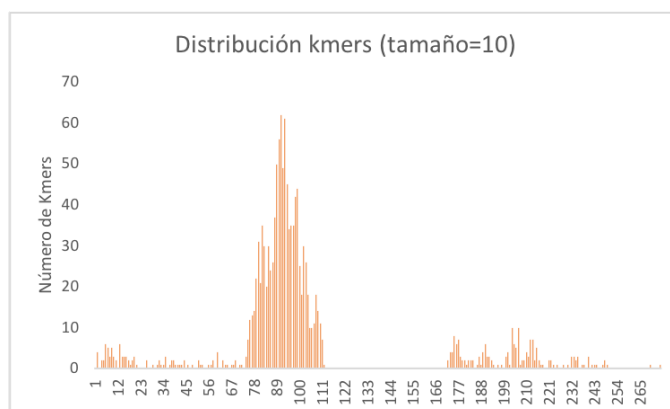


Figura 25: Distribución de k-mers de tamaño 10 para la muestra *test100x*.

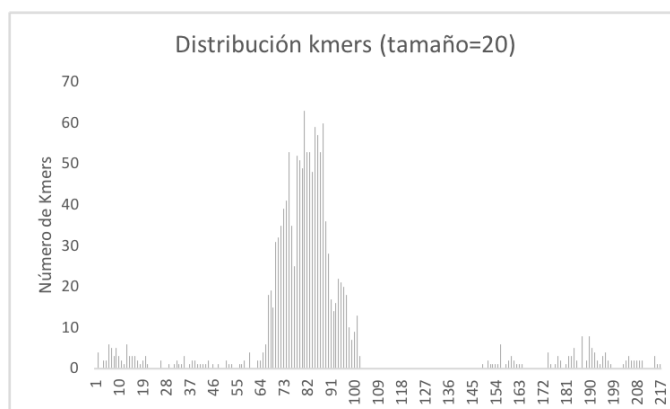


Figura 26: Distribución de k-mers de tamaño 20 para la muestra *test100x*.

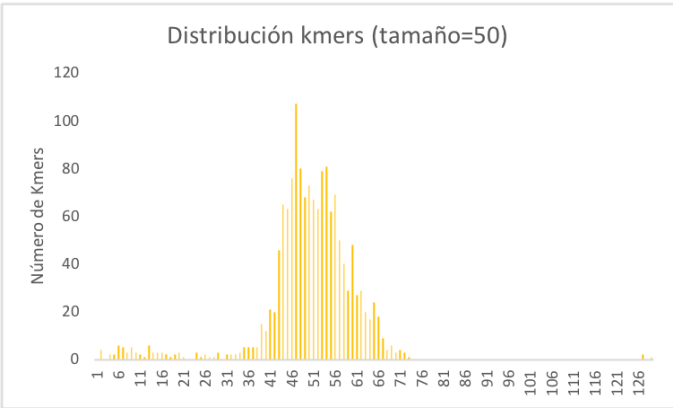


Figura 27: Distribución de k-mers de tamaño 50 para la muestra *test100x*.

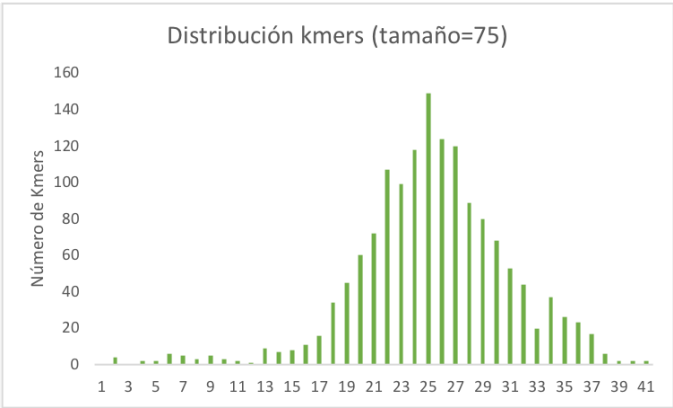


Figura 28: Distribución de k-mers de tamaño 75 para la muestra *test100x*.

Cuadro 4: Resultados de las pruebas realizadas variando el tamaño de los k-mers generados.

	5	10	20	50	75
Número de k-mers totales	1059	1319	1392	1499	1481
Tiempo (ms)	94	93	112	96	111
Abundancia media	95	91	81	47	25

3. [20 %] Ensamblaje de una secuencia

Al igual que en las secciones pasadas, se realizaron varias pruebas para observar el comportamiento del algoritmo. En un principio, se obtuvieron los resultados para cada secuencia en cuanto a la secuencia ensamblada y su tiempo de ejecución. Adicionalmente, se varió el parámetro del solapamiento mínimo para ver si hay un cambio considerable en la secuencia ensamblada. Los valores sobre los cuales se probaron fueron 10, 20, 50 y 95.

3.1. test10x

3.1.1. Solapamiento = 10

El tiempo de ensamblaje de la secuencia en este caso fue de 3ms, y la cadena obtenida fue:

lada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazon la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oido. De otro. Sera de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escr

3.1.2. Solapamiento = 20

El tiempo de ensamblaje de la secuencia en este caso fue de 4ms, y la cadena obtenida fue:

lada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazon la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oido. De otro. Sera de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escr

3.1.3. Solapamiento = 50

El tiempo de ensamblaje de la secuencia en este caso fue de 2ms, y la cadena obtenida fue:

tes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca

3.1.4. Solapamiento = 95

El tiempo de ensamblaje de la secuencia en este caso fue de 0ms, y la cadena obtenida fue vacia.

3.2. test20x

3.2.1. Solapamiento = 10

El tiempo de ensamblaje de la secuencia en este caso fue de 9ms, y la cadena obtenida fue:

ltimo poema de Pablo Neruda. Puedo escribir los versos mas tristes esta noche. Escribir, por ejemplo: La noche esta estrellada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y

canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escr

3.2.2. Sobrelape = 20

El tiempo de ensamblaje de la secuencia en este caso fue de 10ms, y la cadena obtenida fue:

ltimo poema de Pablo Neruda. Puedo escribir los versos mas tristes esta noche. Escribir, por ejemplo: La noche esta estrellada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escr

3.2.3. Sobrelape = 50

El tiempo de ensamblaje de la secuencia en este caso fue de 4ms, y la cadena obtenida fue:

ltimo poema de Pablo Neruda. Puedo escribir los versos mas tristes esta noche. Escribir, por ejemplo: La noche esta estrellada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea

el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escr

3.2.4. Sobrelape = 95

El tiempo de ensamblaje de la secuencia en este caso fue de 3ms, y la cadena obtenida fue vacia.

3.3. test50x

3.3.1. Sobrelape = 10

El tiempo de ensamblaje de la secuencia en este caso fue de 20ms, y la cadena obtenida fue:

lada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escribo.

3.3.2. Sobrelape = 20

El tiempo de ensamblaje de la secuencia en este caso fue de 25ms, y la cadena obtenida fue:

lada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos arboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escribo.

3.3.3. Sobrelape = 50

El tiempo de ensamblaje de la secuencia en este caso fue de 17ms, y la cadena obtenida fue:

lada, y tiritan, azules, los astros, a lo lejos. El viento de la noche gira en el cielo y canta. Puedo escribir los versos mas tristes esta noche. Yo la quise, y a veces ella tambien me quiso. En las noches como esta la tuve entre mis brazos. La bese tantas veces bajo el cielo infinito. Ella me quiso, a veces yo tambien la queria. Como no haber amado sus grandes ojos fijos. Puedo escribir los versos mas tristes esta noche. Pensar que no la tengo. Sentir que la he perdido. Oir la noche inmensa, mas inmensa sin

ella. Y el verso cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos árboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el último dolor que ella me causa, y estos sean los últimos versos que yo le escribo.

3.3.4. Sobrelape = 95

El tiempo de ensamblaje de la secuencia en este caso fue de 5ms, y la cadena obtenida fue:

l cielo infinito. Ella me quiso, a veces yo también la quería. Como no haber amado sus grandes ojos fijos. Puedo esc

3.4. test100x

3.4.1. Sobrelape = 10

El tiempo de ensamblaje de la secuencia en este caso fue de 43ms, y la cadena obtenida fue:

lada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos árboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el último dolor que ella me causa, y estos sean los últimos versos que yo le escribo

3.4.2. Sobrelape = 20

El tiempo de ensamblaje de la secuencia en este caso fue de 59ms, y la cadena obtenida fue:

lada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos árboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el último dolor que ella me causa, y estos sean los últimos versos que yo le escribo

3.4.3. Sobrelape = 50

El tiempo de ensamblaje de la secuencia en este caso fue de 43ms, y la cadena obtenida fue:

lada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos. Mi alma no se contenta con haberla perdido. Como para acercarla mi mirada la busca. Mi corazón la busca, y ella no esta conmigo. La misma noche que hace blanquear los mismos árboles. Nosotros, los de entonces, ya no somos los mismos. Ya no la quiero, es cierto, pero cuanto la quise. Mi voz buscaba el viento para tocar su oído. De otro. Será de otro. Como antes de mis besos. Su voz, su cuerpo claro. Sus ojos infinitos. Ya no la quiero, es cierto, pero tal vez la quiero. Es tan corto el amor, y es tan largo el olvido. Porque en noches como

esta la tuve entre mis brazos, mi alma no se contenta con haberla perdido. Aunque este sea el ultimo dolor que ella me causa, y estos sean los ultimos versos que yo le escribo

3.4.4. Sobrelape = 95

El tiempo de ensamblaje de la secuencia en este caso fue de 2ms, y la cadena obtenida fue:

cae al alma como al pasto el rocío. Que importa que mi amor no pudiera guardarla. La noche esta estrellada y ella no esta conmigo. Eso es todo. A lo lejos alguien canta. A lo lejos.

A partir de los resultados obtenidos con las secuencias de prueba, se puede determinar que entre mayor sea el sobrelape mínimo permitido la secuencia final va a ser más reducida. Esto se debe a que los arcos que se agregan entre cada par de vertices del grafo es menor puesto que no se cumple la condición del sobrelape mínimo. En algunos casos, al poner un número muy alto, va a impedir que el algoritmo realice de forma efectiva el ensamblaje de la secuencia hasta tal punto que no encuentra un par de cadenas que cumplan la restricción.

4. [20 %] Generación Aleatoria de Secuencias

Se generaron secuencias aleatorias del primer fago secuenciado en la historia, Phi174 utilizando tamaños de secuencia de 50, 100, 200, 500 y con profundidad promedio de 5X, 10X, 20X, 50X y 100X. La tabla 6 muestra los resultados obtenidos en los tiempos de generación del grafo y tiempo de ensamblaje para el comando *Overlap*. Adicionalmente, la tabla 5 tiene los resultados obtenidos usando el comando *kmers*.

Cuadro 5: Resultados de las pruebas realizadas utilizando las secuencias aleatorias generadas del fago Phi174 con el comando *Kmers*.

		Tiempo (ms)
len50	5x	79
	10x	55
	20x	132
	50x	134
	100x	177
len100	5x	148
	10x	126
	20x	183
	50x	300
	100x	432
len200	5x	235
	10x	281
	20x	398
	50x	789
	100x	1327
len500	5x	610
	10x	905
	20x	1971
	50x	3860
	100x	8071

Cabe resaltar que se hicieron las pruebas corriendo el programa utilizando 4Gb de memoria RAM, lo cual ocasiona cambiar los parámetros de entrada del algoritmo.

Cuadro 6: Resultados de las pruebas realizadas utilizando las secuencias aleatorias generadas del fago Phi174 con el comando *Overlap*.

len50	5x	tiempo grafo (ms)	58
		tiempo ensamble (ms)	5
	10x	tiempo grafo (ms)	95
		tiempo ensamble (ms)	5
	20x	tiempo grafo (ms)	268
		tiempo ensamble (ms)	9
	50x	tiempo grafo (ms)	1163
		tiempo ensamble (ms)	23
len100	100x	tiempo grafo (ms)	4081
		tiempo ensamble (ms)	28
	5x	tiempo grafo (ms)	137
		tiempo ensamble (ms)	8
	10x	tiempo grafo (ms)	405
		tiempo ensamble (ms)	8
	20x	tiempo grafo (ms)	1311
		tiempo ensamble (ms)	19
len200	50x	tiempo grafo (ms)	7745
		tiempo ensamble (ms)	40
	100x	tiempo grafo (ms)	27999
		tiempo ensamble (ms)	67
	5x	tiempo grafo (ms)	748
		tiempo ensamble (ms)	14
	10x	tiempo grafo (ms)	2699
		tiempo ensamble (ms)	25
len500	20x	tiempo grafo (ms)	10187
		tiempo ensamble (ms)	41
	50x	tiempo grafo (ms)	56391
		tiempo ensamble (ms)	72
	100x	tiempo grafo (ms)	209024
		tiempo ensamble (ms)	146
	5x	tiempo grafo (ms)	532
		tiempo ensamble (ms)	53
len1000	10x	tiempo grafo (ms)	146
		tiempo ensamble (ms)	62
	20x	tiempo grafo (ms)	28396
		tiempo ensamble (ms)	132
	50x	tiempo grafo (ms)	638598
		tiempo ensamble (ms)	134
	100x	tiempo grafo (ms)	1737786
		tiempo ensamble (ms)	532

5. [10 %] Bono

Se utilizó la secuencia del genoma de la bacteria SRR7772462. Debido al tamaño del archivo *fastq* no se logró realizar el ensamblaje por recursos computacionales. Pese a esto, se encontró la distribución de kmers. La figura 29, que tuvo un total de 13984 kmers distintos y solo se muestran la distribución para los primeros 253.

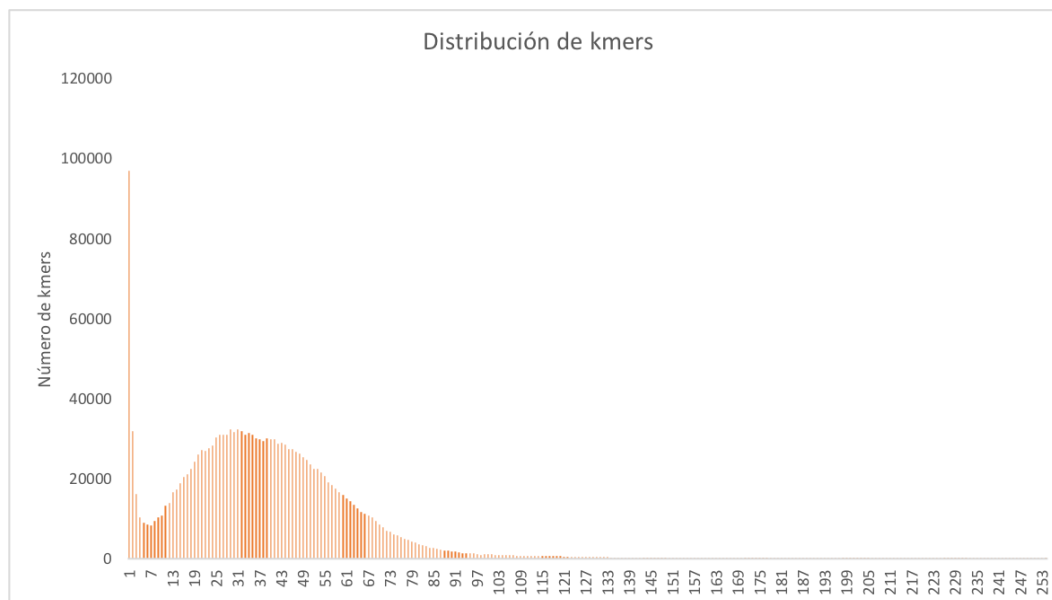


Figura 29: Distribución de k-mers de tamaño 75 para el genoma de la bacteria SRR7772462.

6. Código

Toda la implementación realizada se encuentra en el siguiente repositorio,
[https://github.com/steff456/ComputationalBiology/tree/master/Tarea %202/ReadsAnalyzer](https://github.com/steff456/ComputationalBiology/tree/master/Tarea%202/ReadsAnalyzer)