

Złożenie genomu pszenicy zwyczajnej, odmiany Euphoria (*Triticum aestivum* v. *Euphoria*)
adnotacja oraz analiza porównawcza z dostępnymi genomami

M. Hryc



Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Cele

Cele

- 1 Złożenie genomu *Triticum Aestivum* v. *Euphoria*
- 2 Adnotacja genomu
- 3 Ocena jakości złożenia i adnotacji
- 4 Porównanie otrzymanego złożenia z innymi odmianami pszenicy

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

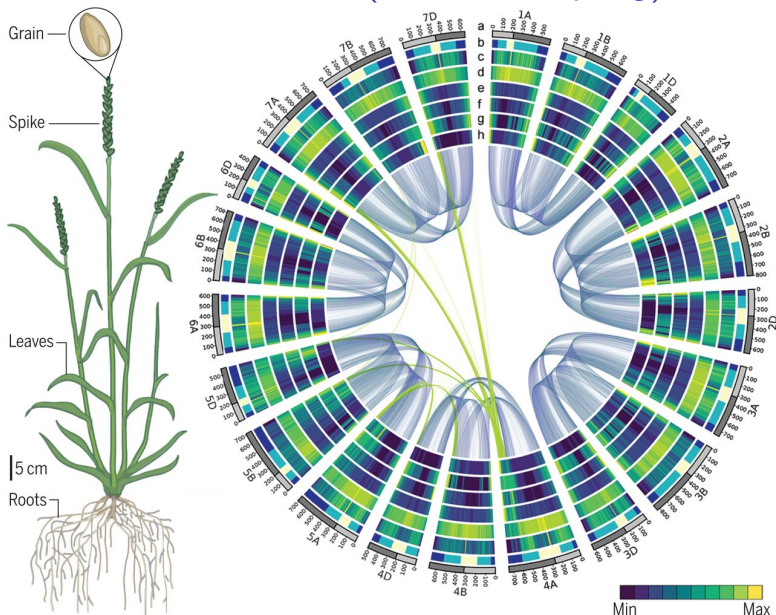
Dyskusja

Stan Wiedzy i innowacyjność

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Genom IWGSC (v. *Chinese Spring*)



Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Table 2. Relative proportions of the major elements of the wheat genome. Proportions of TEs are given as the percentage of sequences assigned to each superfamily relative to genome size. Abbreviations in parentheses under the headings "Class 1" and "Class 2" indicate transposon types.

Major elements	Wheat subgenome			
	AA	BB	DD	Total
Assembled sequence assigned to chromosomes (Gb)	4.935	5.180	3.951	14.066
Size of TE-related sequences (Gb)	4.240	4.388	3.285	11.913
TEs (%)	85.9	84.7	83.1	84.7
Class 1				
LTR-retrotransposons				
Gypsy (RLG)	50.8	46.8	41.4	46.7
Copia (RLC)	17.4	16.2	16.3	16.7
Unclassified LTR-retrotransposons (RLX)	2.6	3.5	3.7	3.2
Non-LTR-retrotransposons				
Long interspersed nuclear elements (RIX)	0.81	0.96	0.93	0.90
Short interspersed nuclear elements (SIX)	0.01	0.01	0.01	0.01
Class 2				
DNA transposons				
CACTA (DTC)	12.8	15.5	19.0	15.5
Mutator (DTM)	0.30	0.38	0.48	0.38
Unclassified with terminal inverted repeats	0.21	0.20	0.22	0.21
Harbinger (DTH)	0.15	0.16	0.18	0.16
Mariner (DTT)	0.14	0.16	0.17	0.16
Unclassified class 2	0.05	0.08	0.05	0.06
hAT (DTA)	0.01	0.01	0.01	0.01
Helitrons (DHH)	0.0046	0.0044	0.0036	0.0042
Unclassified repeats	0.55	0.85	0.63	0.68
Coding DNA	0.89	0.89	1.11	0.95
Unannotated DNA	13.2	14.4	15.7	14.4
(Pre)-microRNAs	0.039	0.057	0.046	0.047
tRNAs	0.0056	0.0050	0.0068	0.0057

Euforia

EUFORIA jest zimotrwałą odmianą pszenicy ozimej otrzymanej bez GMO, więc obecnie bez poznanego genomu.

Inne odmiany, ze złożeniem dobrej jakości

Alchemy, Arinalrfor, Cadenza, **Chinese Spring**, Claire, Jagger, Julius, Karięga, Lancer, Landmakr, Mace, Sy Mattis, Norin61, Paragon, Renan, Robigus, Stanley, Weebil.

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Schemat realizacji i narzędzia

Schemat procedury (inspiracja)

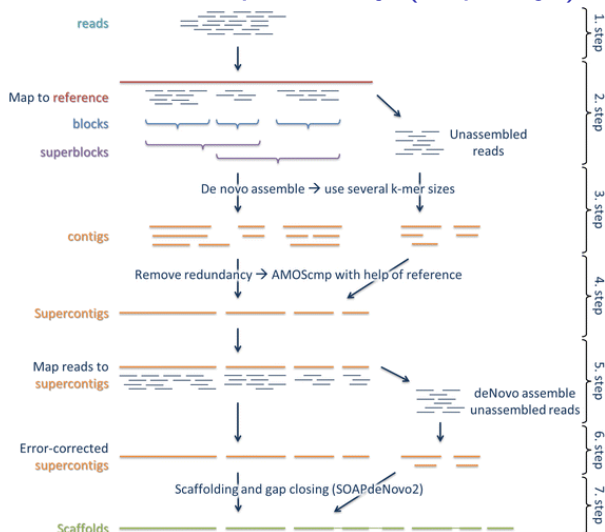


Figure 2: <https://doi.org/10.1186/s12859-017-1911-6>

Sekwencjonowanie

- WGS pbCLR (Continuous Long Reads),
prawdopodobieństwo błędu <20%, większość <10kbp ale
są też <100kbp
- WGS PE Illumina, Q $\approx$ 30, 2×150bp (przed przycinaniem)
- Genom referencyjny: Ensembl Plants release 62 IWGSC
toplevel

Etapy i narzędzia

Wykonane wcześniej

- ① QC i obróbka odczytów:
 - ① PE: FastQC, fastp
 - ② pbCLR: rabbitqcplus, bam2fasta
- ② Składanie *de novo*:
 - ① tylko PE: Abyss
 - ② *tylko long*: wtdbg2, canu, Flye
 - ③ *hybrydowe*: MaSuRCA

Narzędzia, cont.

Składanie wspierane referencją:

- ① Mapowanie PE: BWA-mem2 (do całej referencji); output: BAM
- ② Mapowanie pbCLR: winnowmap2 (do pojedynczych chromosomów); output: PAF
- ③ Blokowanie zmapowań: GNU sort, bedtools, samtools
- ④ Składanie:
 - ① hybrydowe: MaSuRCA
 - ② tylko long: Flye, canu

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Wyniki

Złożenie *de novo* z krótkich odczytów (dla kontekstu)

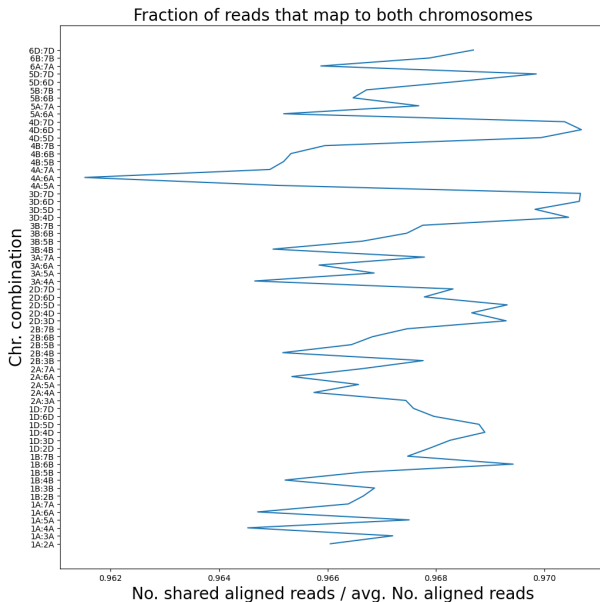
Statistics without reference abyss_2-scaffolds.fa.scaffolds

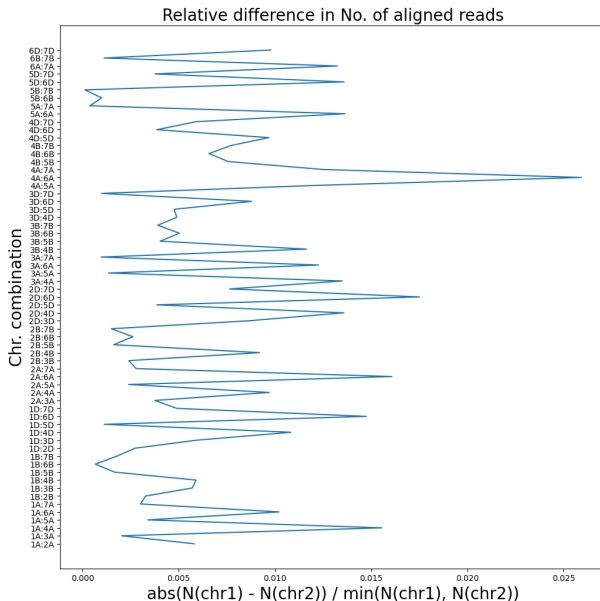
# contigs	4 245 582
# contigs (>= 0 bp)	49 340 672
# contigs (>= 1000 bp)	1 500 504
# contigs (>= 5000 bp)	306 873
# contigs (>= 10000 bp)	66 699
# contigs (>= 25000 bp)	1721
# contigs (>= 50000 bp)	18
Largest contig	109 527
Total length	7 360 414 943
Total length (>= 0 bp)	15 597 369 119
Total length (>= 1000 bp)	5 521 064 809
Total length (>= 5000 bp)	2 565 859 777
Total length (>= 10000 bp)	924 748 898
Total length (>= 25000 bp)	51 742 806
Total length (>= 50000 bp)	1 054 906
N50	3222
N90	629
auN	4819.3
L50	586 226
L90	2 925 345
GC (%)	47.17

Per base quality

# N's per 100 kbp	210.86
# N's	15 520 408

Niespecyficzność zmapowań





Cont.

Do każdego chromosomu zmapowało się 55-59 % wszystkich odczytów

Składanie chromosomu 1A...

JobID	JobName	Partition	State	Elapsed	ReqMe	All	Req
36473538	01_asm.sh	proxima-c+	RUNNING	3-22:27:51	180G	40	40
36473538.ba+	batch		RUNNING	3-22:27:51		40	40
36473538.ex+	extern		RUNNING	3-22:27:51		40	40
36476037	canu01	proxima-c+	RUNNING	3-03:22:43	180G	40	40
36476037.ba+	batch		RUNNING	3-03:22:43		40	40
36476037.ex+	extern		RUNNING	3-03:22:43		40	40
36481024	flye01	proxima-c+	RUNNING	2-00:30:38	180G	40	40
36481024.ba+	batch		RUNNING	2-00:30:38		40	40
36481024.ex+	extern		RUNNING	2-00:30:38		40	40

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Dyskusja

Główne trudności

- ① Mało informatywne zmapowania
- ② Rozmiar problemu:
 - ① Duży genom (~15Gbp)
 - ② Duże dane (odczyty; 374GB po kompresji gzip)
- ③ Większość genomu pszenicy to sekwencje powtórzone
- ④ Po świętach oczekiwanie na zasoby na klastrze trwa kilkanaście godzin

Wyniki są gorsze niż w znanych mi publikacjach, gdzie stosowano prostsze podejścia, co może sugerować że największym problemem są dane: niskiej jakości długie odczyty, niepełne pokrycie sekwencjonowania, zbyt niska głębokość PE (20-30x).

Złożenie
genomu
pszenicy
zwyczajnej,
odmiany
Euphoria
(*Triticum
aestivum* v.
Euphoria)

M. Hryc

Cele

Stan Wiedzy i
innowacyjność

Schemat
realizacji i
narzędzia

Wyniki

Dyskusja

Pszennica ozima	
Pszonżyto ozime	
Pszonżyto ozime	
Jęczmień ozimy	