



GENOMOK ÉS BÖNGÉSZŐK

Megmutatni a láthatatlant?

2019.10.24

Biológia – megfigyelésen alapuló tudomány

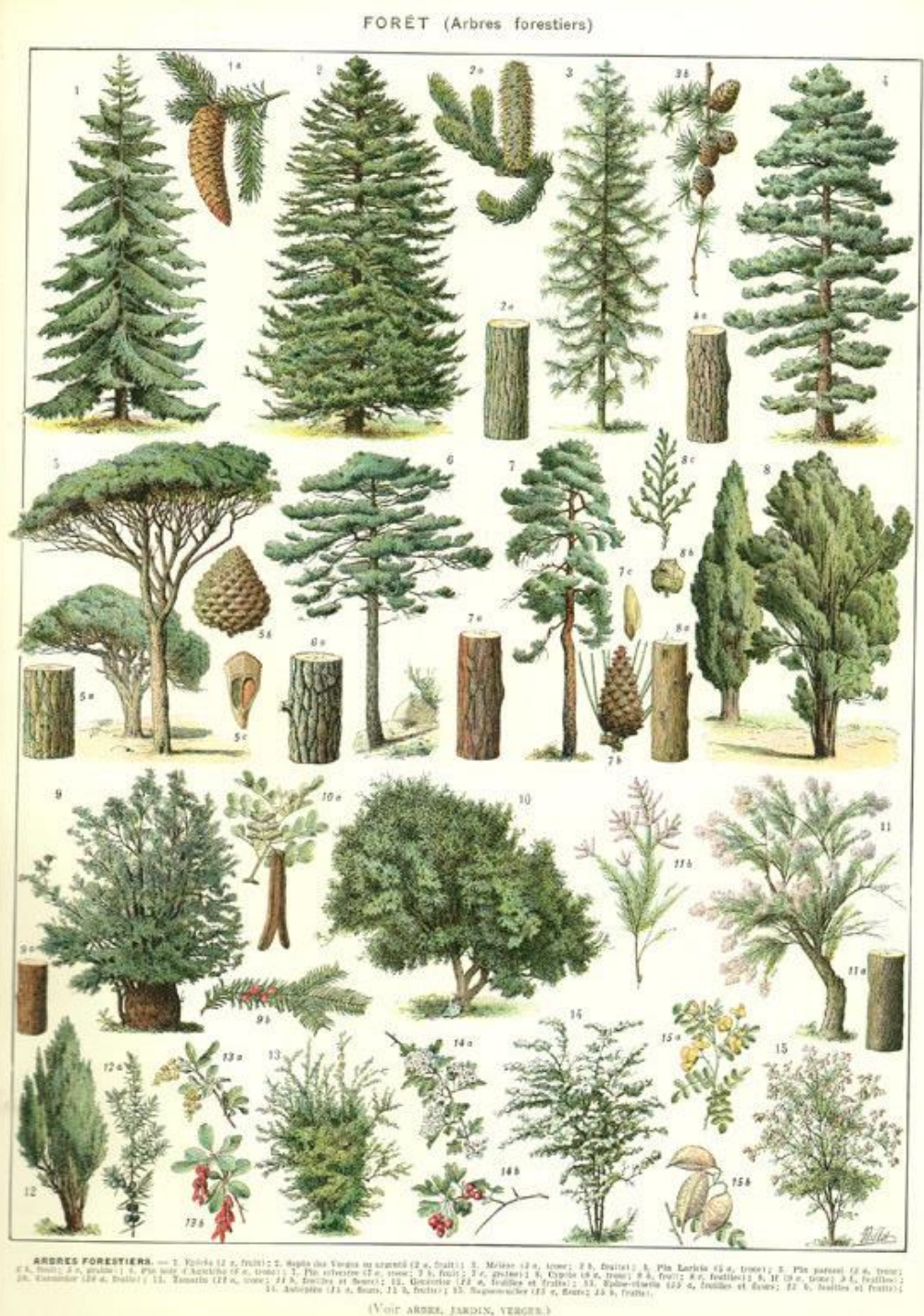


ÁBRÁZOLÁS

A biológiát könnyebb képekben elmesélni, mint szavakkal

Erdei fák

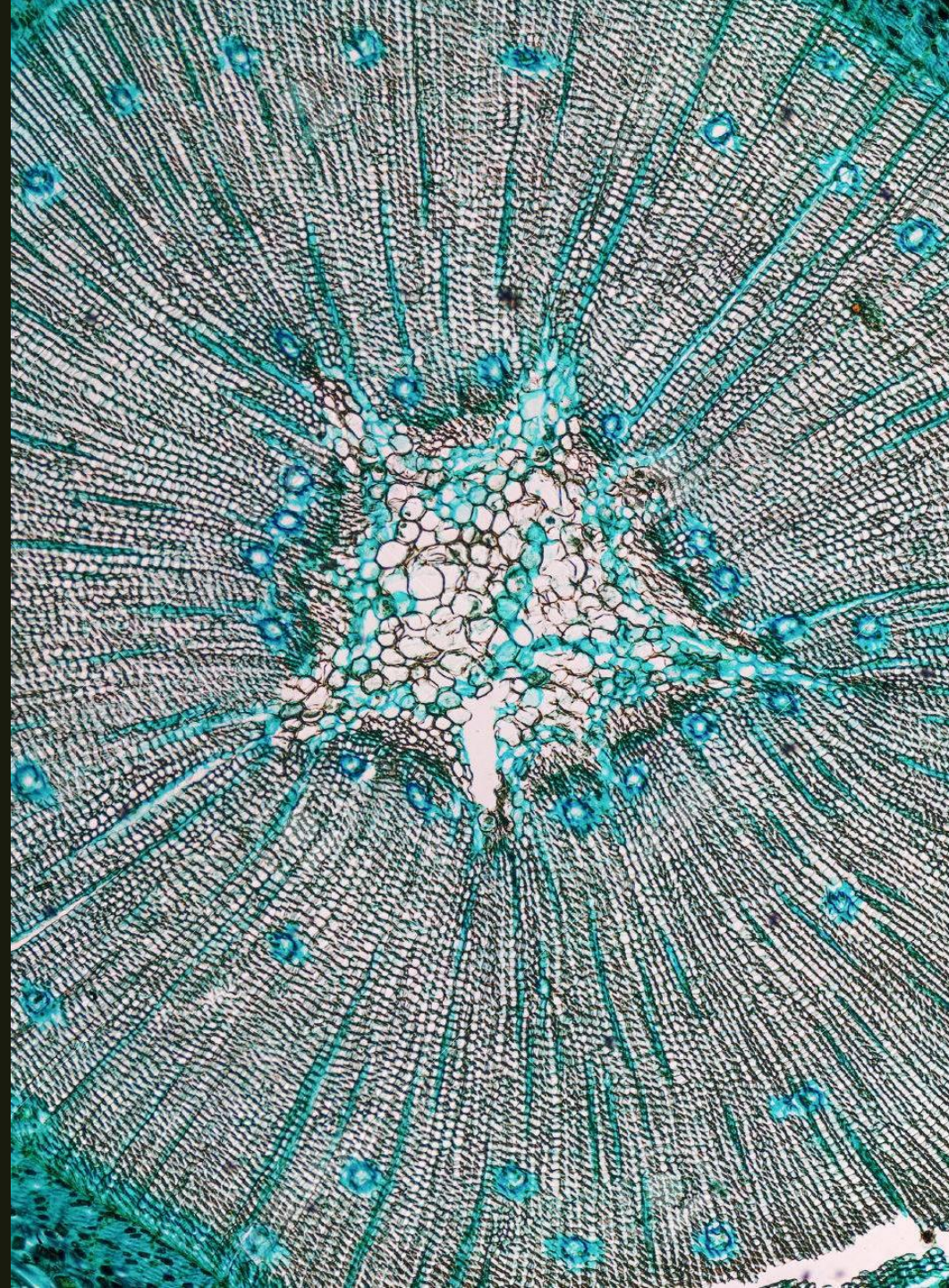
Kézi ábrázolás kora



Fénymikroszkóp – fenyőág metszet

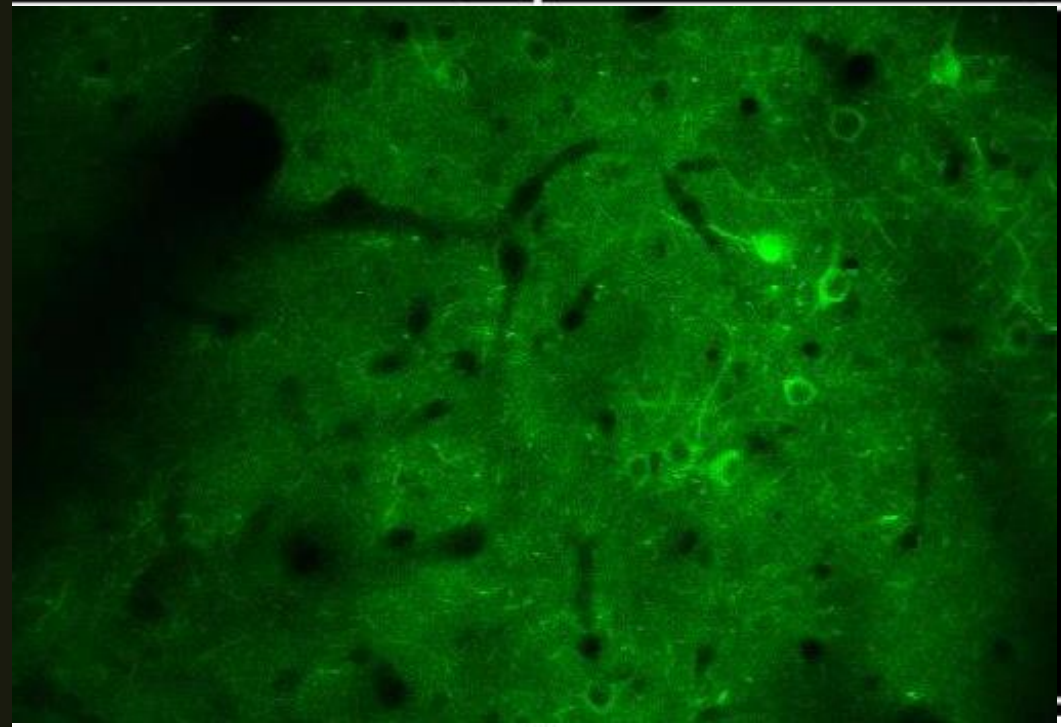
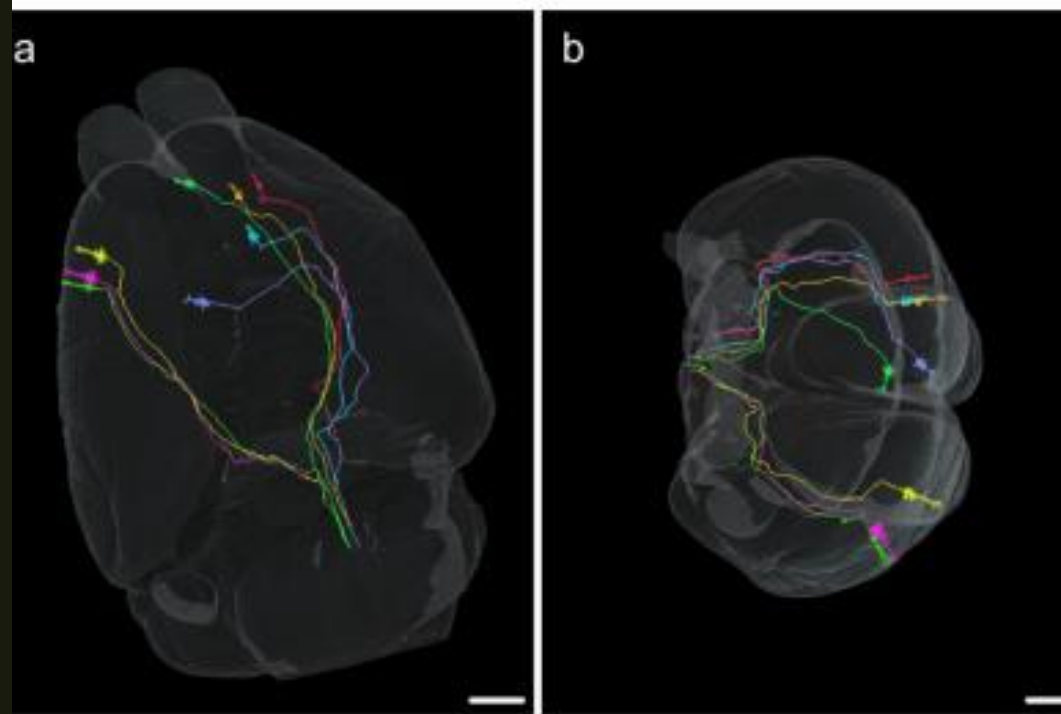
https://kr.123rf.com/photo_34048275_%ED%98%84%EB%AF%B8%EA%B2%BD%EC%9D%84-%ED%86%B5%ED%95%B4-%EB%B3%B8-%EC%86%8C%EB%82%98%EB%AC%B4-%EB%82%98%EB%AC%B4-%EB%82%98%EB%AC%B4-%EB%8B%A8%EB%A9%B4%EC%9D%98-%EA%B3%A0%ED%95%B4%EC%83%81%EB%8F%84-%EB%B9%9B-photomicrograph.html?fromid=anRnbyt2dDVoM29rcVMxQndFUhd1dz09

Szabolcs Makai. All rights reserved. Not for sharing.



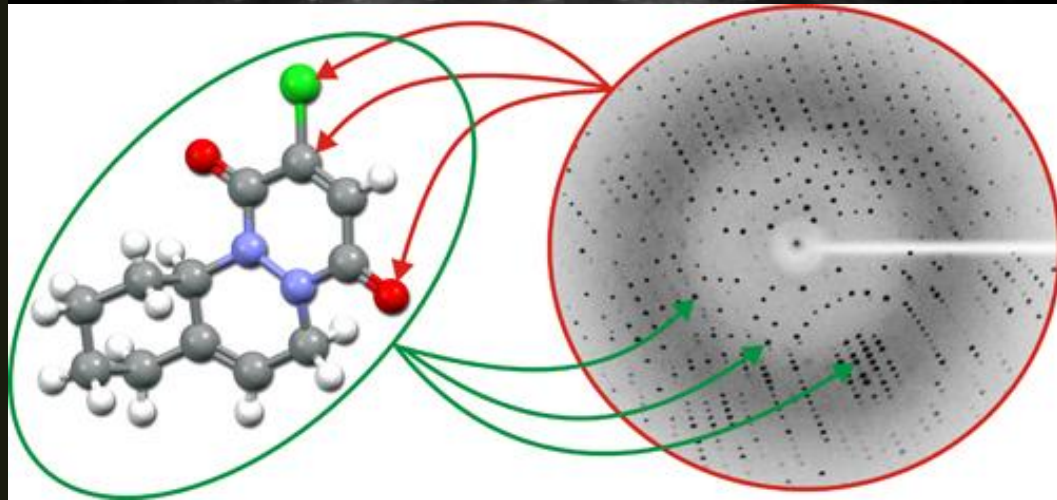
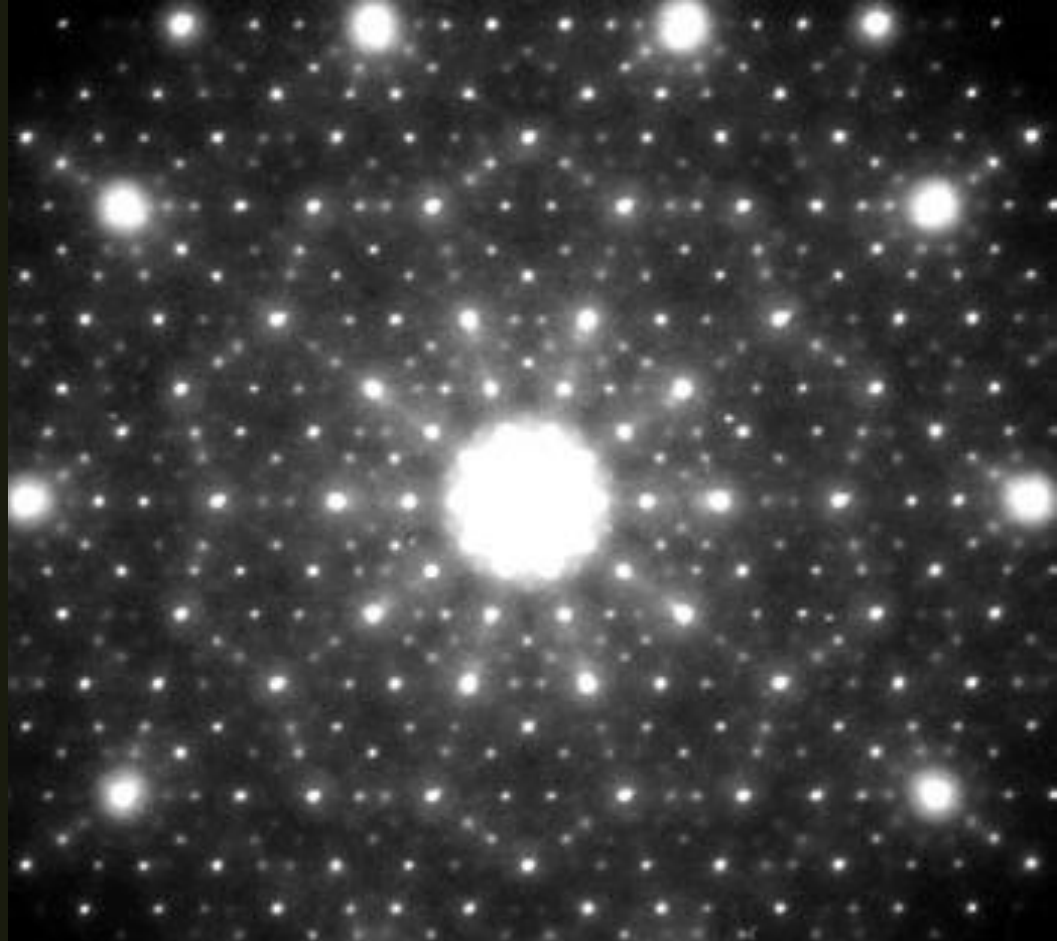
Kétfoton mikroszkópia

SOTE



Röntgen diffrakció

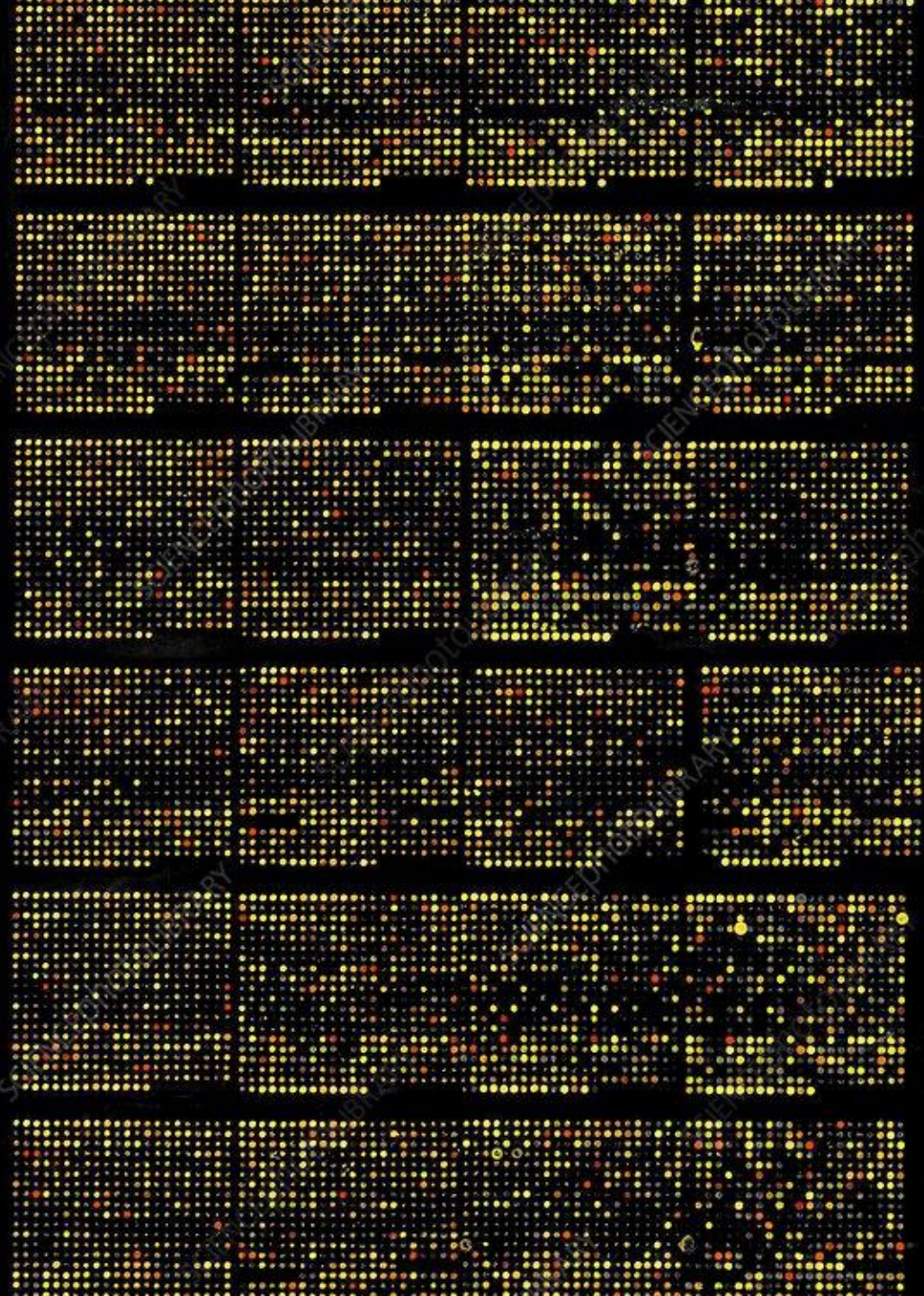
kristályszerkezet



Microarray – génexpresszió

[https://www.sciencephoto.com
/media/629130/view/dna-
microarrays](https://www.sciencephoto.com/media/629130/view/dna-microarrays)

Szabolcs Makai. All rights reserved. Not for sharing.



BIG DATA



MEETS BIOLOGY

MEGJELENÍTÉS

Lerajzolni a láthatatlant

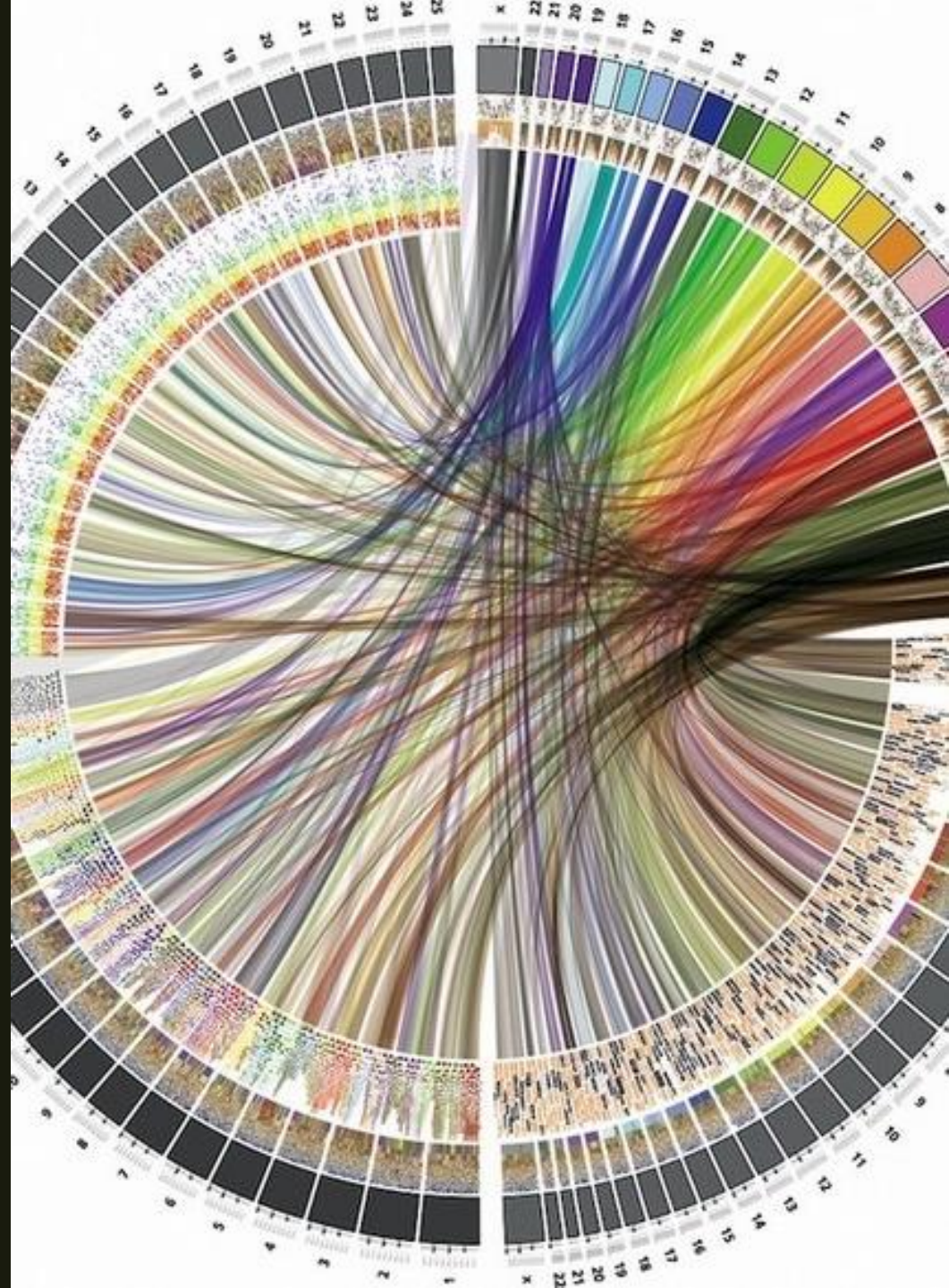
Mit?

- Molekulák
- Adatok és adatállományok
- Eredmények, értékelések
- Bármit, amit hatékonyan akarunk kommunikálni

Genetika rokonság: csimpánz és ember

<https://www.wired.com/2013/11/wired-data-life-martin-krzywinski>

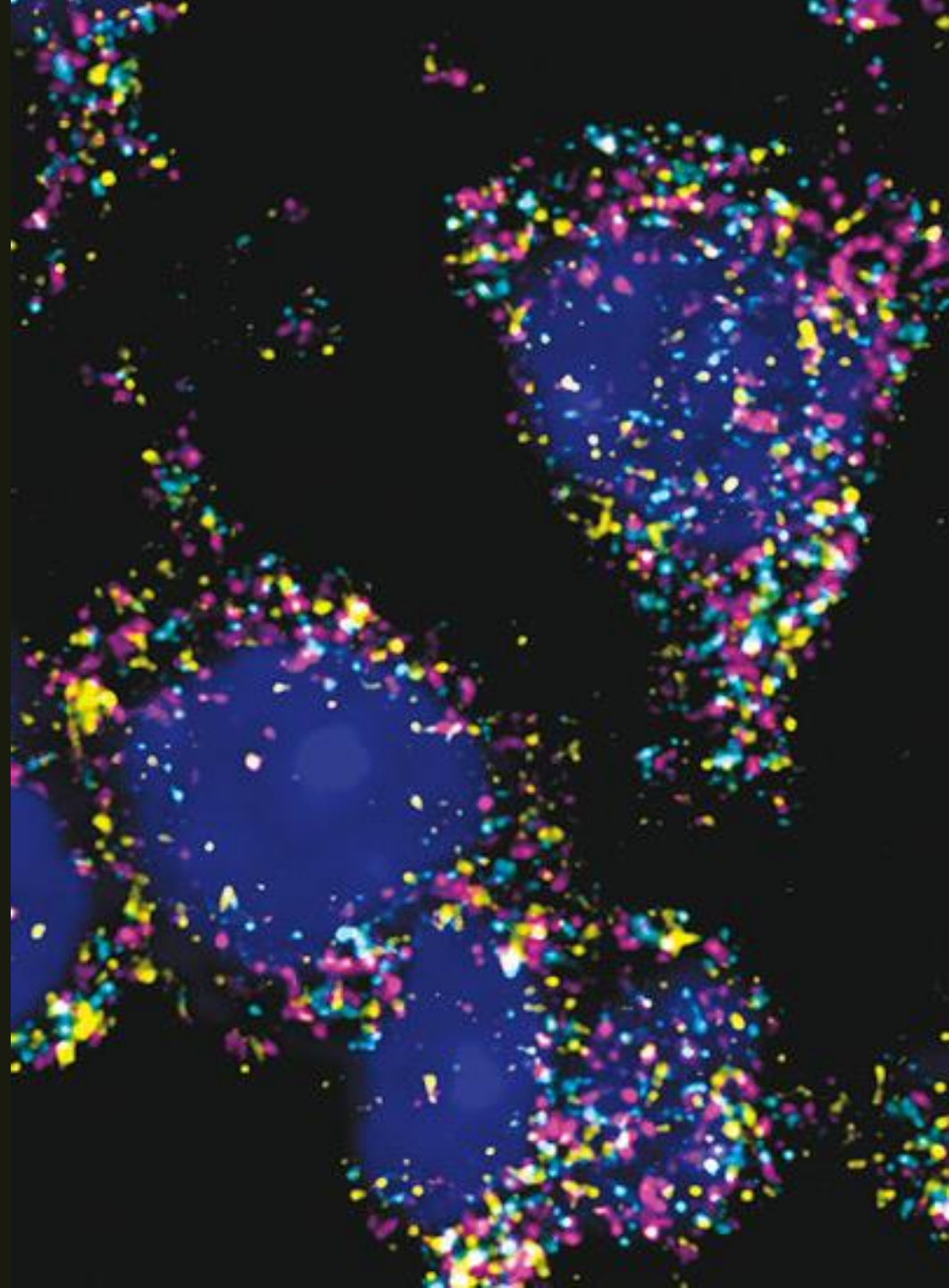
Szabolcs Makai. All rights reserved. Not for sharing.



A newly developed technique from the laboratory of Long Cai colorfully illuminates every mRNA in every cell within a tissue sample with “super-resolution.” The technique can be applied to study everything from embryos to cancers.

<https://magazine.caltech.edu/post/biology-and-big-data>

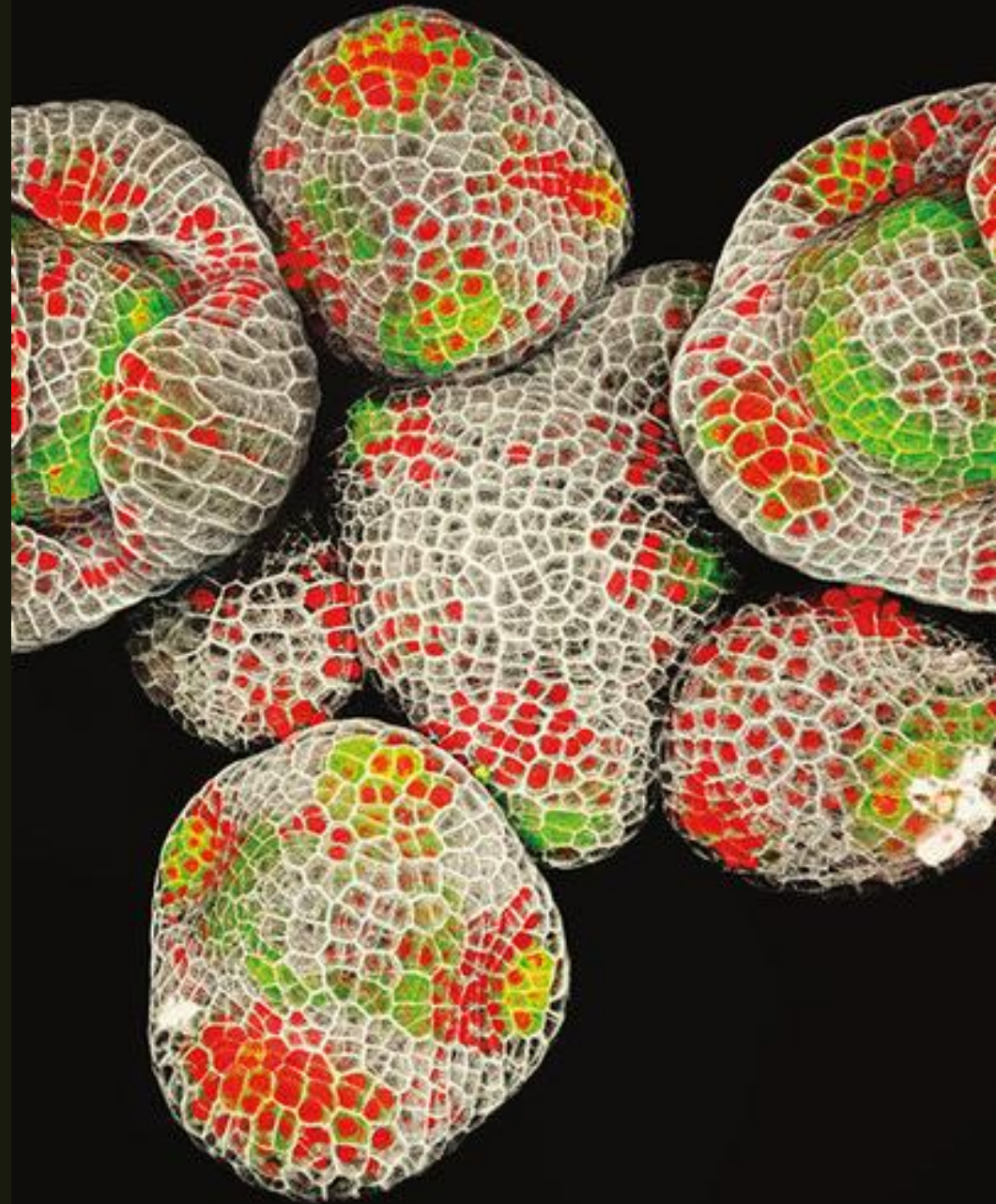
Szabolcs Makai. All rights reserved. Not for sharing.



Ha gépi tanulási módszerek és a megfigyelt adatok találkoznak?

[https://magazine.caltech.edu/
post/biology-and-big-data](https://magazine.caltech.edu/post/biology-and-big-data)

Szabolcs Makai. All rights reserved. Not for sharing.

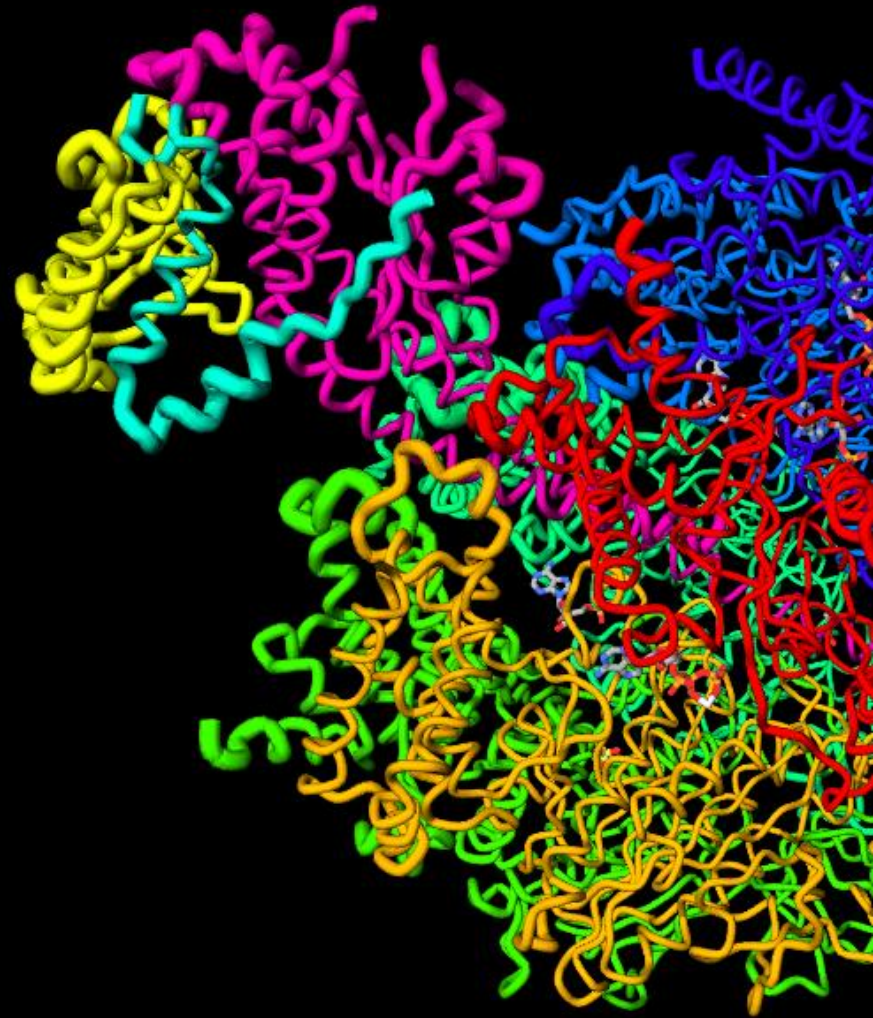


Online molecular viewer

<http://molview.org/?pdbid=1e79>

Szabolcs Makai. All rights reserved. Not for sharing.

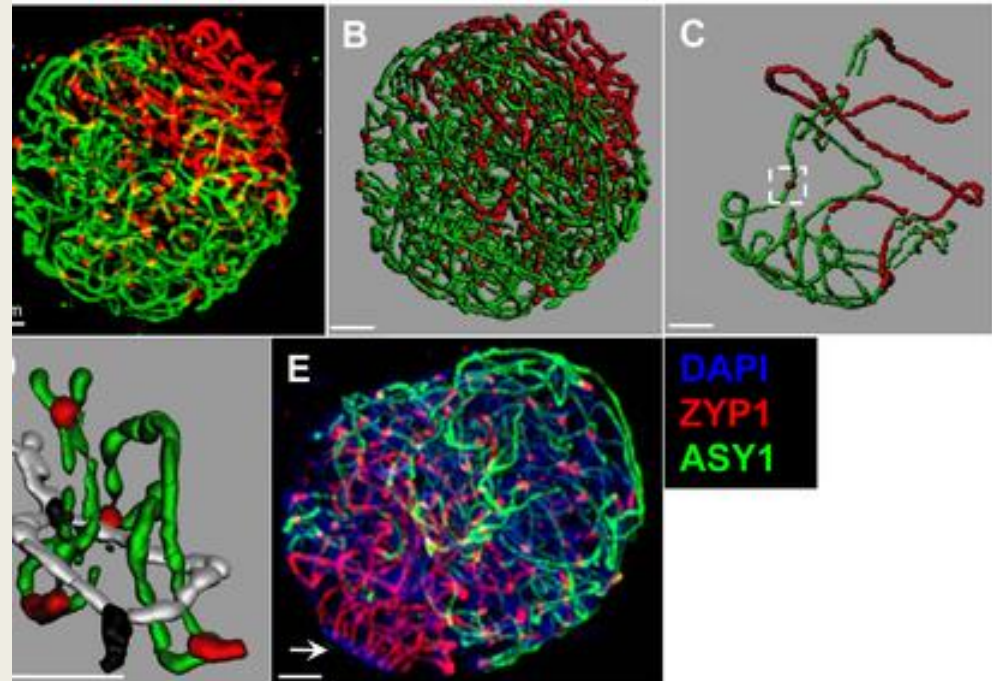
Q ATP Synthase MolView Tools Model Protein



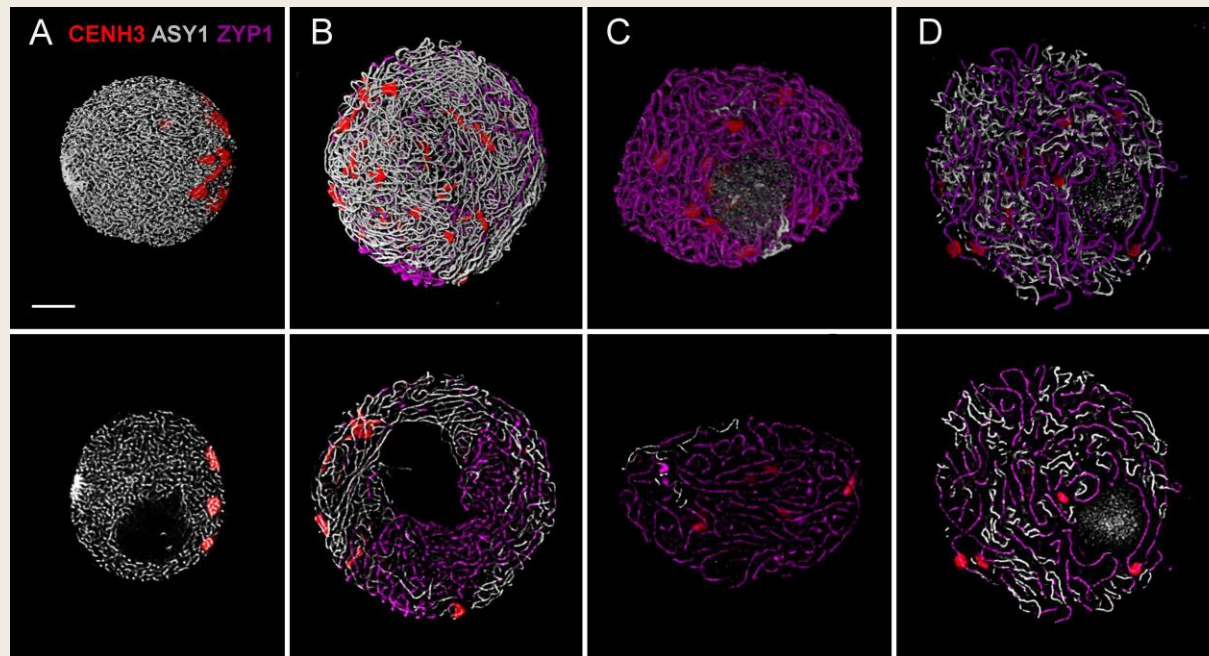
Analysis of synaptic progression during barley prophase I.

C) A single bivalent extracted from the reconstruction in Figure 2B, containing an example of a ZYP1 focus present on a single AE (white box). (D) Detail of an interlock isolated from the nucleus shown in Figure 2A. For ease of interpretation, one of the bivalents has been re-coloured to show ASY1 in grey and ZYP1 in black.

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0039539>



Three-colour immunolabelling on meiotic prophase I (leptotene to diplotene) nuclei of wheat. CENH3 centromeric protein is visualised by a rabbit anti-CENH3 antibody and is shown in red, SC axial element protein ASY1 is visualised by a guinea pig anti-ASY1 antibody and is shown in grey and SC central element protein is visualised by a rat anti-ZYP1 antibody and is pseudo-coloured in purple. Three-dimensionally reconstructed confocal image stacks are shown in the top row while a single z stack of the image is presented in the bottom row. Images show substages of meiotic prophase I: (A) leptotene, (B) mid-zygotene, (C) pachytene, (D) diplotene. Bar = 5 μ m.



ImmunoFISH: Simultaneous Visualisation of Proteins and DNA Sequences Gives Insight Into Meiotic Processes in Nuclei of Grasses

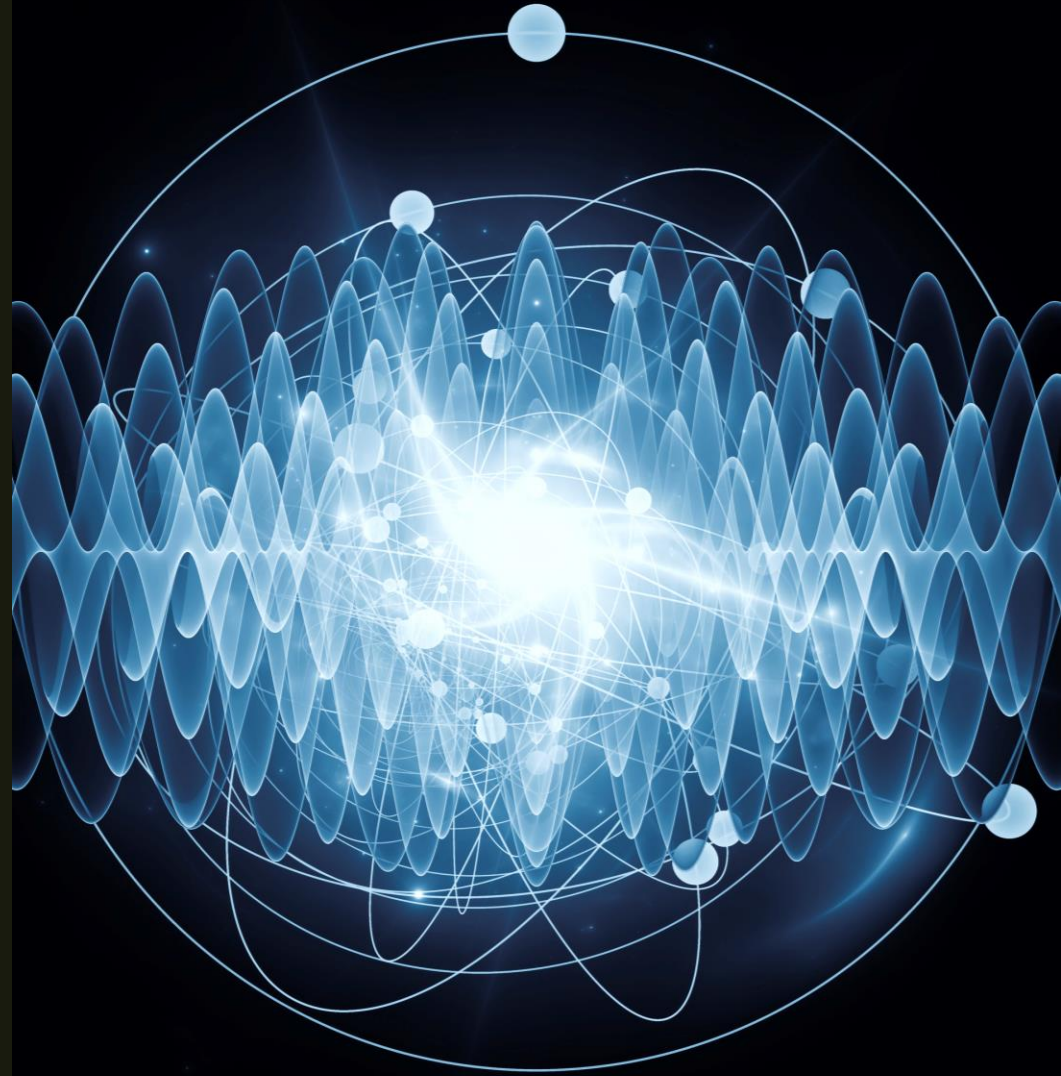
Adél Seps¹, Attila Fáb¹, Katalin Jäger¹, J. S. Heslop-Harrison² and Trude Schwarzacher²

ESZKÖZÖK

Megértés útján

Abstract visualization

Just like quantum physics



Search: for

 e.g. [Carboxy*](#) or [chx28](#)

All genomes

- [View full list of all Ensembl Plants species](#)
- [Edit your favourites](#)

Favourite genomes


Arabidopsis thaliana

TAIR10


Oryza sativa Japonica Group

IRGSP-1.0


Triticum aestivum

IWGSC


Hordeum vulgare

IBSC_v2


Zea mays

B73_RefGen_v4


Physcomitrella patens

Phypa_V3

New wheat assembly

This release of Ensembl Plants hosts the [latest wheat assembly](#) from the IWGSC (RefSeq v1.0). Current data includes:

- The IWGSC RefSeq v1.1 gene annotation.
- Alignment of 98,270 high confidence genes from the TGACv1 annotation.
- The Axiom 820K SNP Array from CerealsDB.
- EMS-induced mutations from sequenced TILLING populations (Kronos and Cadensa).
- Assembly to assembly mapping and gene ID mapping to the previous TGAC v1 assembly and annotation.
- Whole genome alignments to rice, brachypodium and barley.
- The TGACv1 assembly will still be available via the [Ensembl Plants archive site](#).

Polyploid View Enabled

Polyploid view for [wheat](#) has been enabled, allowing users to view alignments between the three wheat components [simultaneously](#).

Funding

Ensembl Plants datasets are constructed in a direct collaboration with the [Gramene](#) resource, funded by the United States [National Science Foundation award #1127112](#). Read [more about our collaboration with Gramene](#).



The development of resources for wheat is funded by the BBSRC-funded [Designing Future Wheat ISP](#).





Triticum aestivum (IWGSC) ▾

Search

e.g. [TraesCS3D02G273600](#) or [3D:2585940-2634711](#) or [Carboxy*](#)

For information about the assembly and annotation please view the [IWGSC announcement](#).

The previous wheat assembly ([TGACv1](#)) and every other plant from release 31 is available in the new [Ensembl Plants archive](#) site.

About *Triticum aestivum*

Triticum aestivum (bread wheat) is a major global cereal grain essential to human nutrition. Wheat was one of the first cereals to be domesticated, originating in the [fertile crescent](#) around 7000 years ago. Bread wheat is hexaploid, with a genome size estimated at ~17 Gbp, composed of three closely-related and independently maintained genomes that are the result of a series of naturally occurring hybridization events. The ancestral progenitor genomes are considered to be [Triticum urartu](#) (the A-genome donor) and an unknown grass thought to be related to [Aegilops speltoides](#) (the B-genome donor). This first hybridization event produced tetraploid emmer wheat (AABB, *T. dicoccoides*) which hybridized again with [Aegilops tauschii](#) (the D-genome donor) to produce modern bread wheat.

Taxonomy ID [4565](#)

Data source [International Wheat Genome Sequencing Consortium](#)



[More information and statistics](#)

Genome assembly: [IWGSC](#)



[More information and statistics](#)



[Download DNA sequence \(FASTA\)](#)



[Convert your data to IWGSC coordinates](#)



[Display your data in Ensembl Plants](#)



[View karyotype](#)



[Example region](#)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.



[More about this genebuild](#)



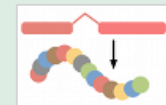
[Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3](#)



[Update your old Ensembl IDs](#)

Carboxy* CAB
RuBisCO
ADH
F-box
PSII

[Example gene](#)



[Example transcript](#)

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.



[More about comparative analyses](#)



[Phylogenetic overview of gene families](#)



[Download alignments \(EMF\)](#)



[Genomic alignments \[5\] \[Show»\]](#)



[Example gene tree](#)

Variation

What can I find? Short sequence variants.



[More about variation in Triticum aestivum](#)



[More about variation in Ensembl Plants](#)



[Download all variants - GVF - VCF - VEP](#)

[Variant Effect Predictor](#)



ATCGAGCT
ATCCAGCT
ATCGAGAT

[Example variant](#)



Genome

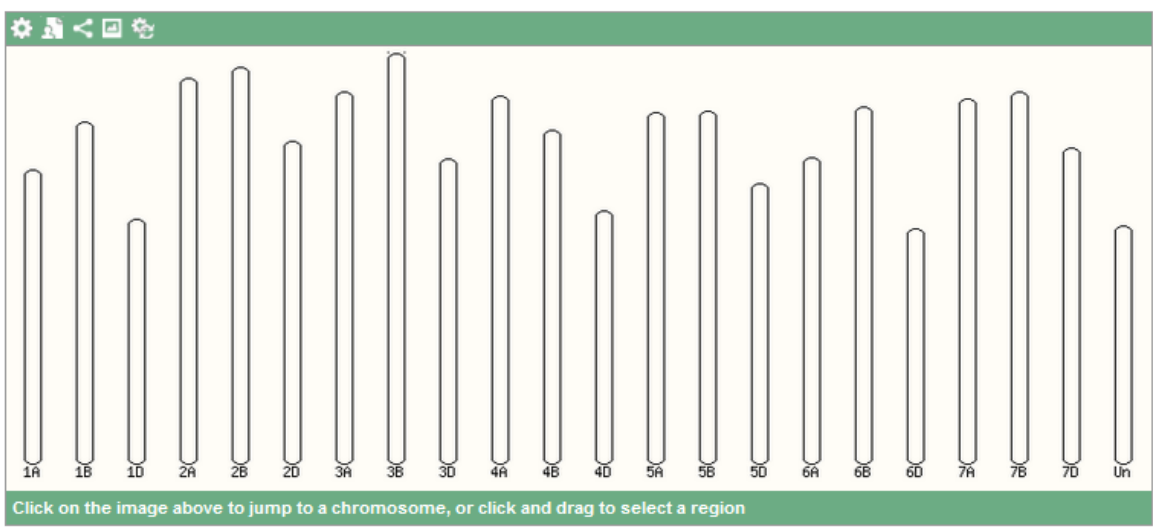
- Location-based displays
- Whole genome
 - Chromosome summary
 - Region overview
 - Region in detail
 - Comparative Genomics
 - Synteny
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
 - Polyploid view
 - Genetic Variation
 - Variant table
 - Resequencing
 - Strain table
 - Linkage Data
 - Markers

- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

Ensembl Plants is produced in collaboration with Gramene

Whole genome ?

+ Add features



Summary

Assembly	IWGSC, INSDC Assembly GCA_900519105.1 , Jul 2018
Database version	94.4
Base Pairs	14,547,261,565
Golden Path Length	14,547,261,565
Genebuild by	IWGSC
Genebuild method	Imported from IWGSC
Data source	International Wheat Genome Sequencing Consortium

Gene counts

Coding genes	107,891
Non coding genes	12,853
Small non coding genes	12,491
Long non coding genes	362
Gene transcripts	146,597

Other

Short Variants	14,142,687
----------------	------------



Triticum aestivum (IWGSC) ▾

Location: 3D:2,585,940-2,634,711

- Location-based displays
 - Whole genome
 - Chromosome summary
 - Region overview
 - Region in detail**
- Comparative Genomics
 - Synteny
 - Alignments (image)
 - Alignments (text)
 - Region Comparison
 - Polyploid view
- Genetic Variation
 - Variant table
 - Resequencing
 - Linkage Data
- Markers

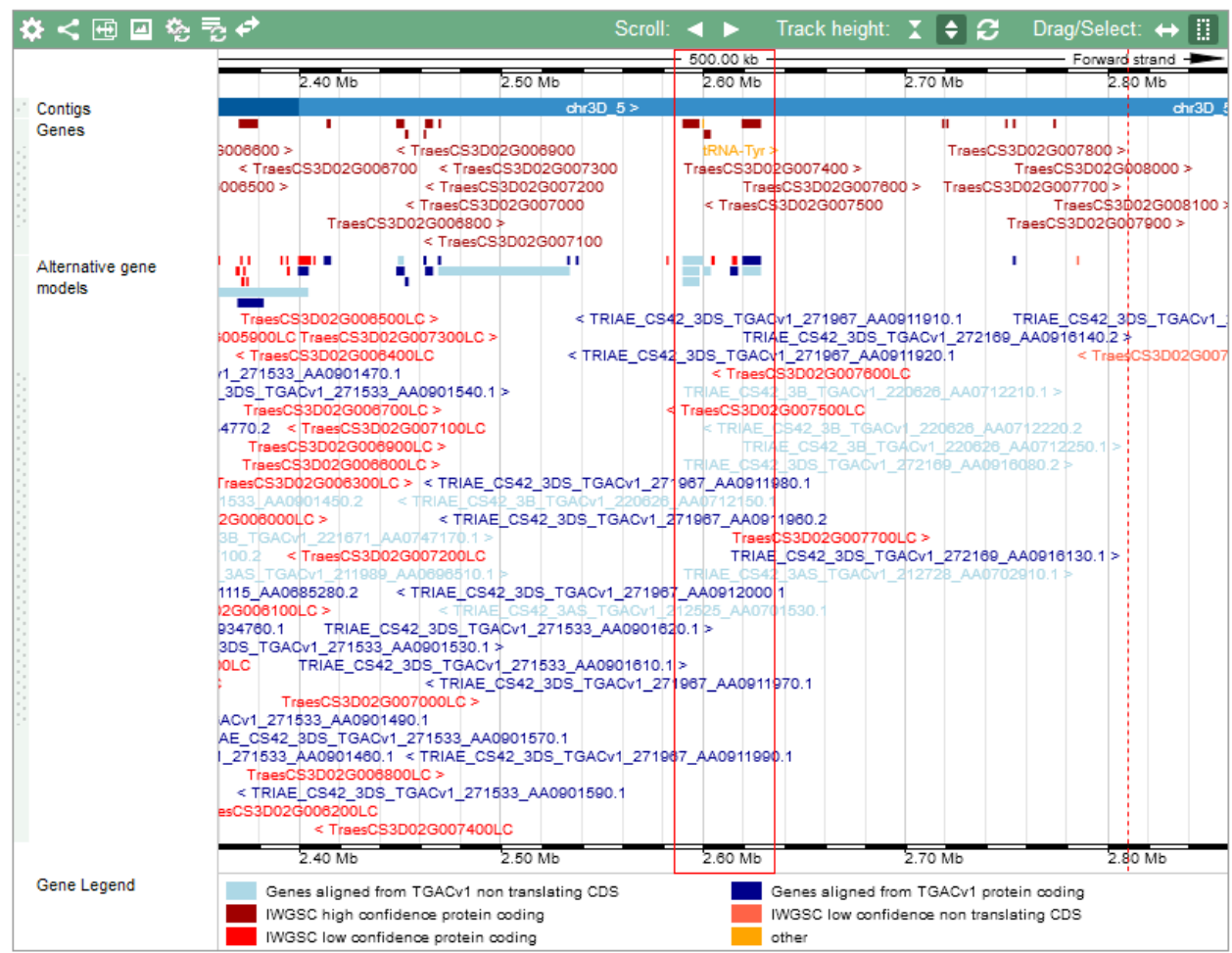
- Configure this page
- Custom tracks
- Export data
- Share this page
- Bookmark this page

[Ensembl Plants is produced in collaboration with Gramene](#)

Chromosome 3D: 2,585,940-2,634,711

Chr. 3D

Region in detail



Location: 3D:2585940-2634711

Go

Gene:

Go



Drag/Select:

Alternative gene models.....

Genes.....

Contigs.....

Genes.....

Alternative gene models.....

Variant - All sources.....

%GC.....

Gene Legend

Protein Coding

- protein coding
- IWGSC low confidence protein coding
- Genes aligned from TGACv1 protein coding

Non-Protein Coding

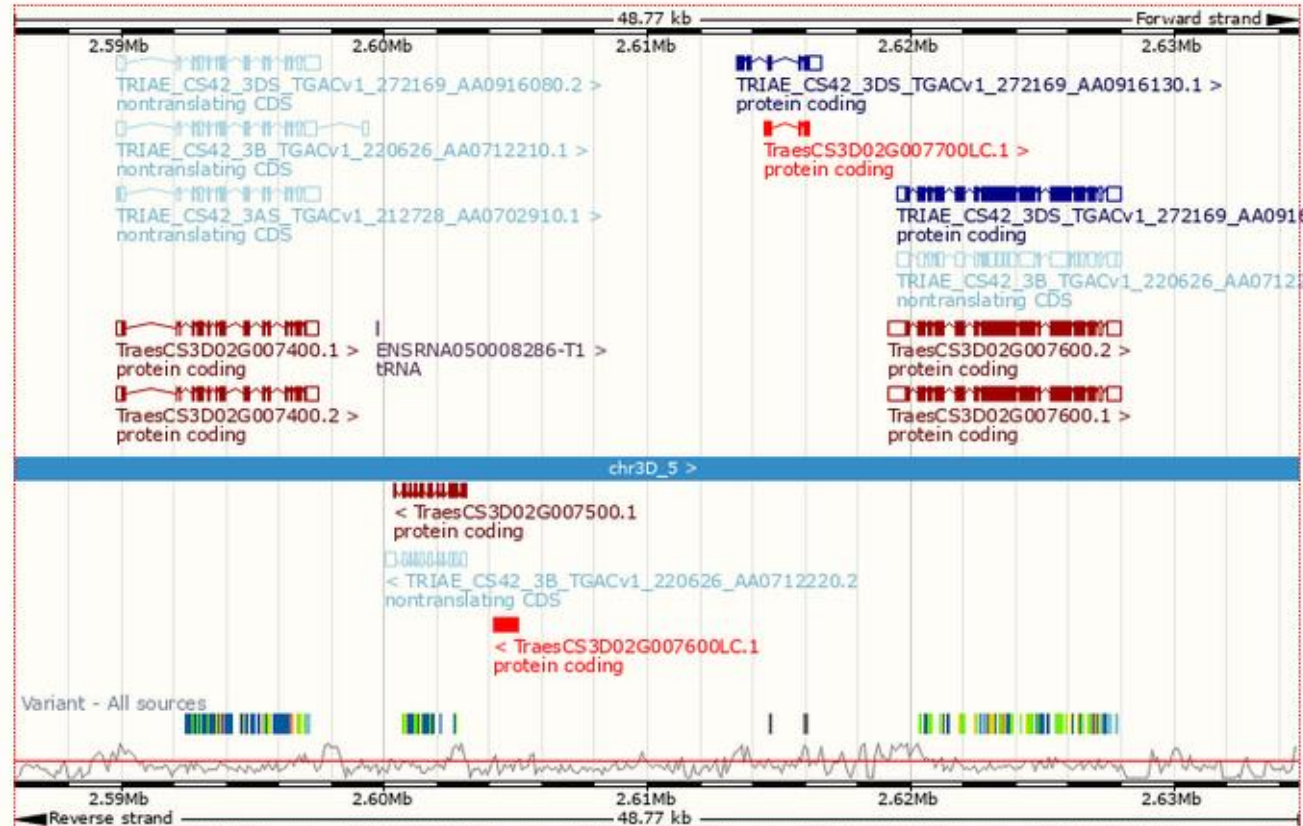
- RNA gene
- Genes aligned from TGACv1 non translating CDS

Variant Legend

- splice donor variant
- missense variant
- synonymous variant
- intron variant
- splice acceptor variant
- splice region variant
- 3 prime UTR variant
- intergenic variant

There are currently 42 tracks turned off.

Ensembl Plants Triticum aestivum version 94.4 (IWGSC) Chromosome 3D: 2,585,940 - 2,634,711



Gene: TraesCS3D02G273600

Description Heat shock protein 101 [Source:UniProtKB/TrEMBL;Acc:Q9SPH4 [↗](#)]

Location [Chromosome 3D: 379,535,906-379,539,827](#) reverse strand.

About this gene This gene has 2 transcripts ([splice variants](#)), [80 orthologues](#) and [21 paralogues](#).

Transcripts [Show transcript table](#)

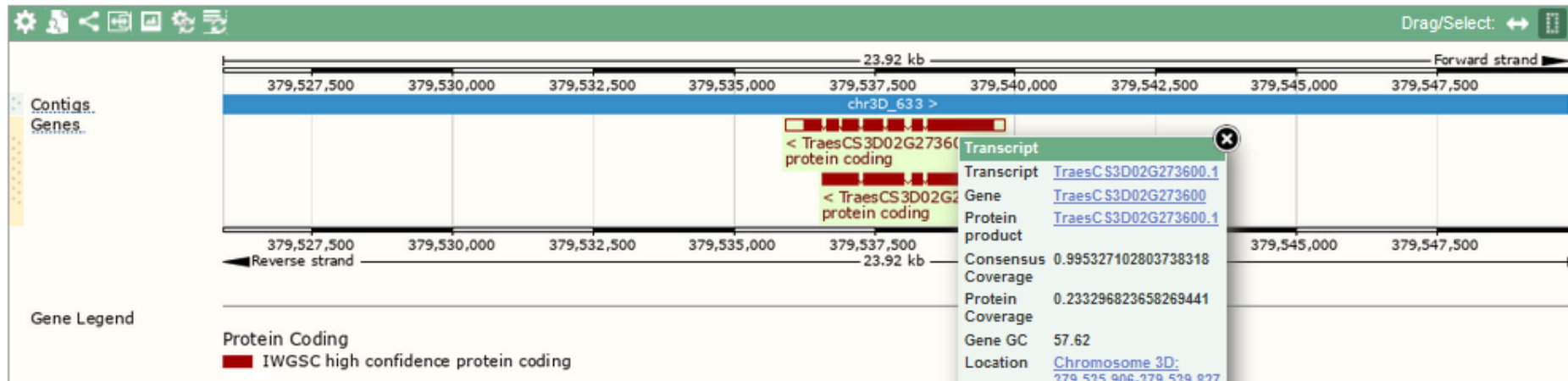
Summary [?](#)

Gene type Protein coding

Annotation method Genes annotated with high confidence by IWGSC



Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)



[?](#) Configuring the display

Tip: use the "Configure this page" link on the left to show additional data in this region.

Gene: TraesCS3D02G273600

Description	Heat shock protein 101 [Source:UniProtKB/TrEMBL;Acc:Q9SPH4 ↗]
Location	Chromosome 3D: 379,535,906-379,539,827 reverse strand.
About this gene	This gene has 2 transcripts (splice variants), 80 orthologues and 21 paralogues .
Transcripts	Show transcript table

GO: Molecular function [?](#)

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs
GO:0000166 ↗	nucleotide binding	IEA	UniProtKB/TrEMBL:Q9SPH4 ↗		TraesCS3D02G273600.1 <ul style="list-style-type: none">Search BioMartView on karyotype
GO:0005524 ↗	ATP binding	IEA	UniProtKB/TrEMBL:Q9SPH4 ↗ , InterPro:ATPase_AAA_core ↗ , InterPro:ClpA/B_CS1 ↗ , InterPro:ClpA/B ↗		TraesCS3D02G273600.2 TraesCS3D02G273600.1 <ul style="list-style-type: none">Search BioMartView on karyotype

Gene: TraesCS3D02G273600

Description	Heat shock protein 101 [Source:UniProtKB/TrEMBL;Acc:Q9SPH4 ↗]
Location	Chromosome 3D: 379,535,906-379,539,827 reverse strand.
About this gene	This gene has 2 transcripts (splice variants), 80 orthologues and 21 paralogues .
Transcripts	Show transcript table

GO: Biological process [?](#)

Accession	Term	Evidence	Annotation source	Mapped using	Transcript IDs
GO:0019538 ↗	protein metabolic process	IEA	UniProtKB/TrEMBL:Q9SPH4 ↗ , InterPro:Clp_N_dom_sf ↗		TraesCS3D02G273600.1 TraesCS3D02G273600.2 <ul style="list-style-type: none">Search BioMartView on karyotype

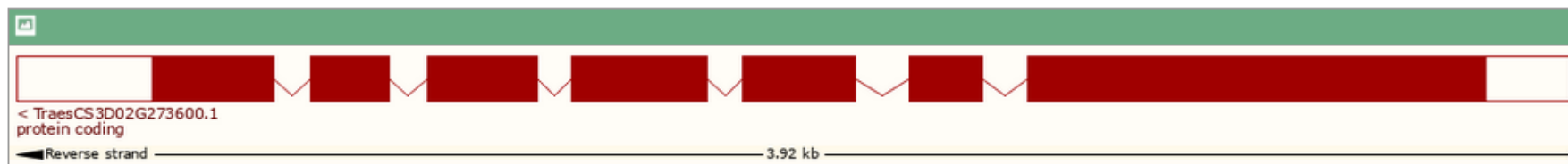
Gene: TraesCS3D02G273600

Trans: TraesCS3D02G273600.1

Transcript: TraesCS3D02G273600.1

Description	Heat shock protein 101 [Source:UniProtKB/TrEMBL;Acc: Q9SPH4]
Location	Chromosome 3D: 379,535,906-379,539,827 reverse strand.
About this transcript	This transcript has 7 exons , is annotated with 32 domains and features and is associated with 130 variations .
Gene	This transcript is a product of gene TraesCS3D02G273600 Show transcript table

Summary



Statistics	Exons: 7, Coding exons: 7, Transcript length: 3,297 bps, Translation length: 913 residues
Version	TraesCS3D02G273600.1.
Type	Protein coding
Annotation Method	Genes annotated with high confidence by IWGSC

Transcript: TraesCS3D02G273600.1

Description Heat shock protein 101 [Source:UniProtKB/TrEMBL;Acc:Q9SPH4]

Location [Chromosome 3D: 379,535,906-379,539,827](#) reverse strand.

About this transcript This transcript has [7 exons](#), is annotated with [32 domains and features](#) and is associated with [130 variations](#).

Gene This transcript is a product of gene [TraesCS3D02G273600](#) [Show transcript table](#)

Domains & features

Domains

Domain source	Start	End	Description	Accession	InterPro
PANTHER	1	859	-	PTHR11638	-
PANTHER	1	859	-	PTHR11638:SF146	-
Gene3D	152	345	-	3.40.50.300	-
CDD	184	344	-	cd00009	-
Gene3D	346	546	-	3.40.50.300	-
Gene3D	552	767	-	3.40.50.300	-
CDD	596	767	-	cd00009	-
Gene3D	768	864	-	1.10.8.60	-
SMART	201	346	AAA+ ATPase domain	SM00382	IPR003593 [Display all genes with this domain]
SMART	599	742	AAA+ ATPase domain	SM00382	IPR003593 [Display all genes with this domain]
Pfam	206	322	ATPase, AAA-type, core	PF00004	IPR003959 [Display all genes with this domain]
Pfam	599	762	ATPase, AAA-type, core	PF07724	IPR003959 [Display all genes with this domain]
Pfam	768	847	Clp ATPase, C-terminal	PF10431	IPR019489 [Display all genes with this domain]
SMART	768	859	Clp ATPase, C-terminal	SM01086	IPR019489 [Display all genes with this domain]
Pfam	17	67	Clp, N-terminal	PF02861	IPR004176 [Display all genes with this domain]
Pfam	98	148	Clp, N-terminal	PF02861	IPR004176 [Display all genes with this domain]
Gene3D	3	151	Clp, N-terminal domain superfamily	1.10.1780.10	IPR036628 [Display all genes with this domain]
Superfamily	4	158	Clp, N-terminal domain superfamily	SSF81923	IPR036628 [Display all genes with this domain]
Prints	603	621	ClpA/B family	PR00300	IPR001270 [Display all genes with this domain]
Prints	648	666	ClpA/B family	PR00300	IPR001270 [Display all genes with this domain]
Prints	677	695	ClpA/B family	PR00300	IPR001270 [Display all genes with this domain]
Prints	710	724	ClpA/B family	PR00300	IPR001270 [Display all genes with this domain]
PROSITE patterns	297	309	ClpA/B, conserved site 1	PS00870	IPR018368 [Display all genes with this domain]
PROSITE patterns	633	651	ClpA/B, conserved site 2	PS00871	IPR028299 [Display all genes with this domain]
Superfamily	164	547	P-loop containing nucleoside triphosphate hydrolase	SSF52540	IPR027417 [Display all genes with this domain]
Superfamily	548	848	P-loop containing nucleoside triphosphate hydrolase	SSF52540	IPR027417 [Display all genes with this domain]



Expression Atlas

Gene expression across
species and biological
conditions

Query single cell expression

[To Single Cell Expression Atlas](#)

[Home](#)

[Browse experiments](#)

[Download](#)

[Release notes](#)

[FAQ](#)


[Help](#)

[Licence](#)

[About](#)

[Also in this section](#)

Exploring gene expression results across species under different biological conditions

Expression Atlas is an open science resource that gives users a powerful way to find information about gene and protein expression across species and biological conditions such as different tissues, cell types, developmental stages and diseases among others. Expression Atlas aims to help answering questions such as 'where is a certain gene expressed?' or 'how does its expression change in a disease?'.


[Read more about Expression Atlas](#)

Search

Gene set enrichment

Gene / Gene properties

Examples: [REG1B](#), [zinc finger](#), [O14777 \(UniProt\)](#), [GO:0010468 \(regulation of gene expression\)](#)

Organism

Any

Biological conditions

Examples: [lung](#), [leaf](#), [valproic acid](#), [cancer](#)

Search

Clear

Results for leaf OR seed AND Triticum aestivum

Baseline expression | Differential expression

Show anatomograms

Filter your results

Triticum aestivum

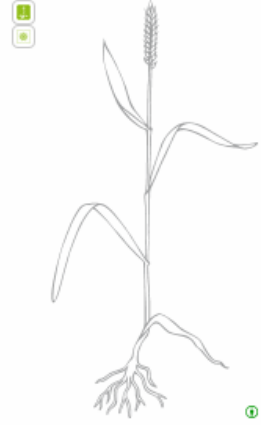
Organism part

Developmental stage

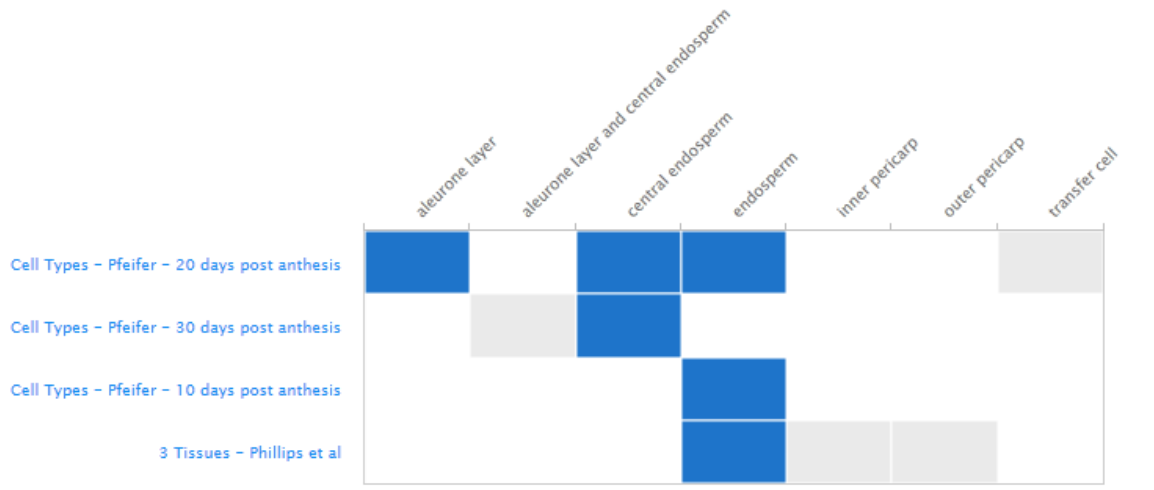
Sampling time point

Organism part

Showing 4 experiments:



By experiment type | Filters | Download



High Medium Low Below cutoff No data available

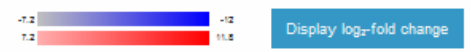
Did you find these results useful?



Results for leaf OR seed AND Triticum aestivum

Baseline expression Differential expression

Filter your results



Download results

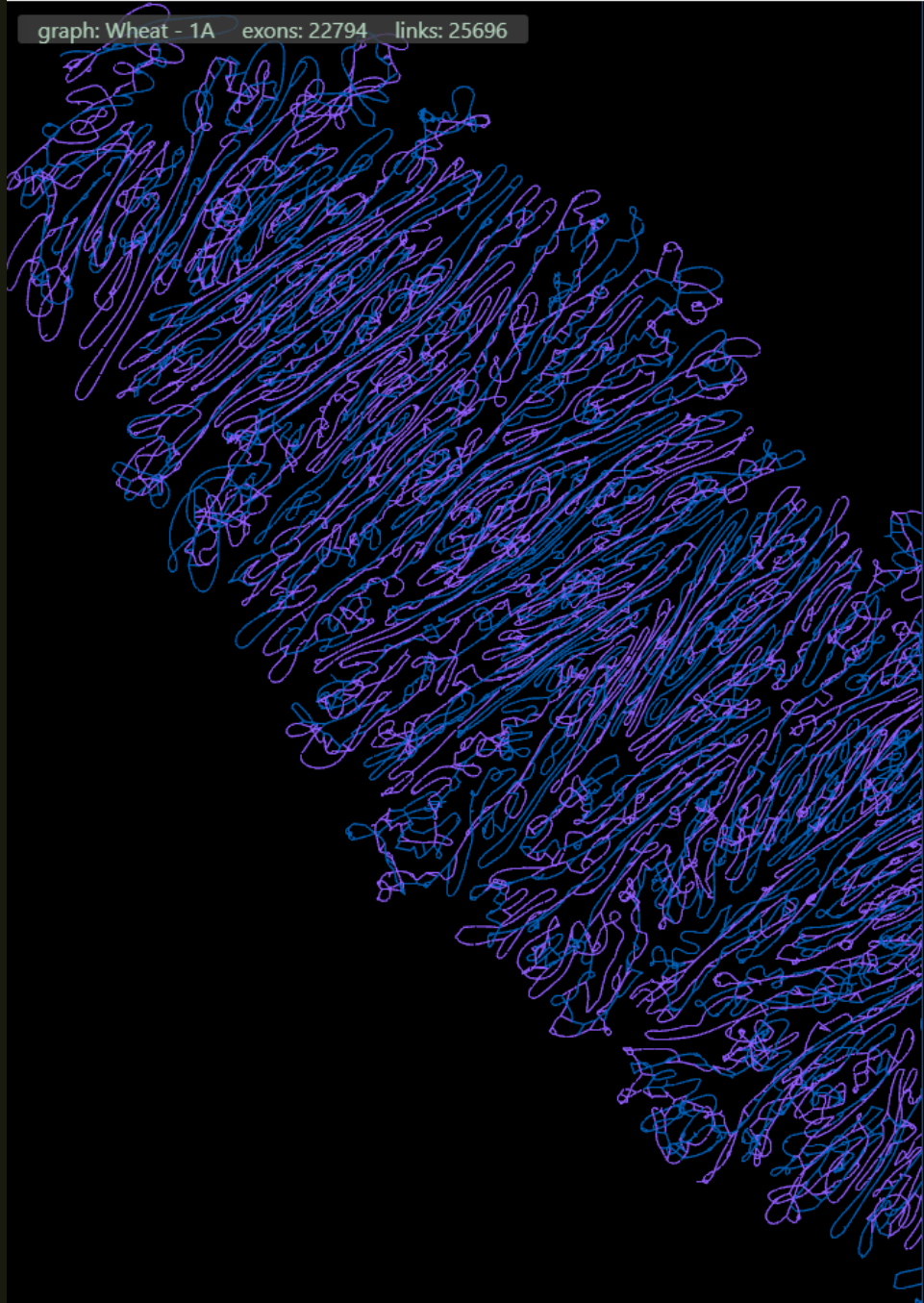
- Kingdom
- Plants
- Species
- Triticum aestivum
- Experiment type
- RNA-seq mRNA differential
- Experimental variables
- Infect
 - Time
 - Organism part
 - Age
 - Environmental stress
- Regulation
- Down
 - Up

Log ₂ -fold change	Species	Gene name	Comparison	Experimental variables	Experiment name
		TRIAE_CS42_5BL_TGACv1_413043_AA1367940	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_2BL_TGACv1_130091_AA0403600	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_2BL_TGACv1_129997_AA0401280	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_2AL_TGACv1_094560_AA0299650	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_7AS_TGACv1_569613_AA1820410	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_5BL_TGACv1_412328_AA1367790	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_4AL_TGACv1_288169_AA0939290	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_2DL_TGACv1_162669_AA0563320	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_5BL_TGACv1_405520_AA1329310	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_5DL_TGACv1_433492_AA1414670	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature
		TRIAE_CS42_7DS_TGACv1_622753_AA2044770	'4 week; cold temperature regimen' vs '2 week; control'	age, environmental stress	Transcription profiling by high throughput sequencing of wheat in response to low temperature



Last opened: D:\bioit\development\RG\rg_wd\wheat_Marzena_v8.6.rag

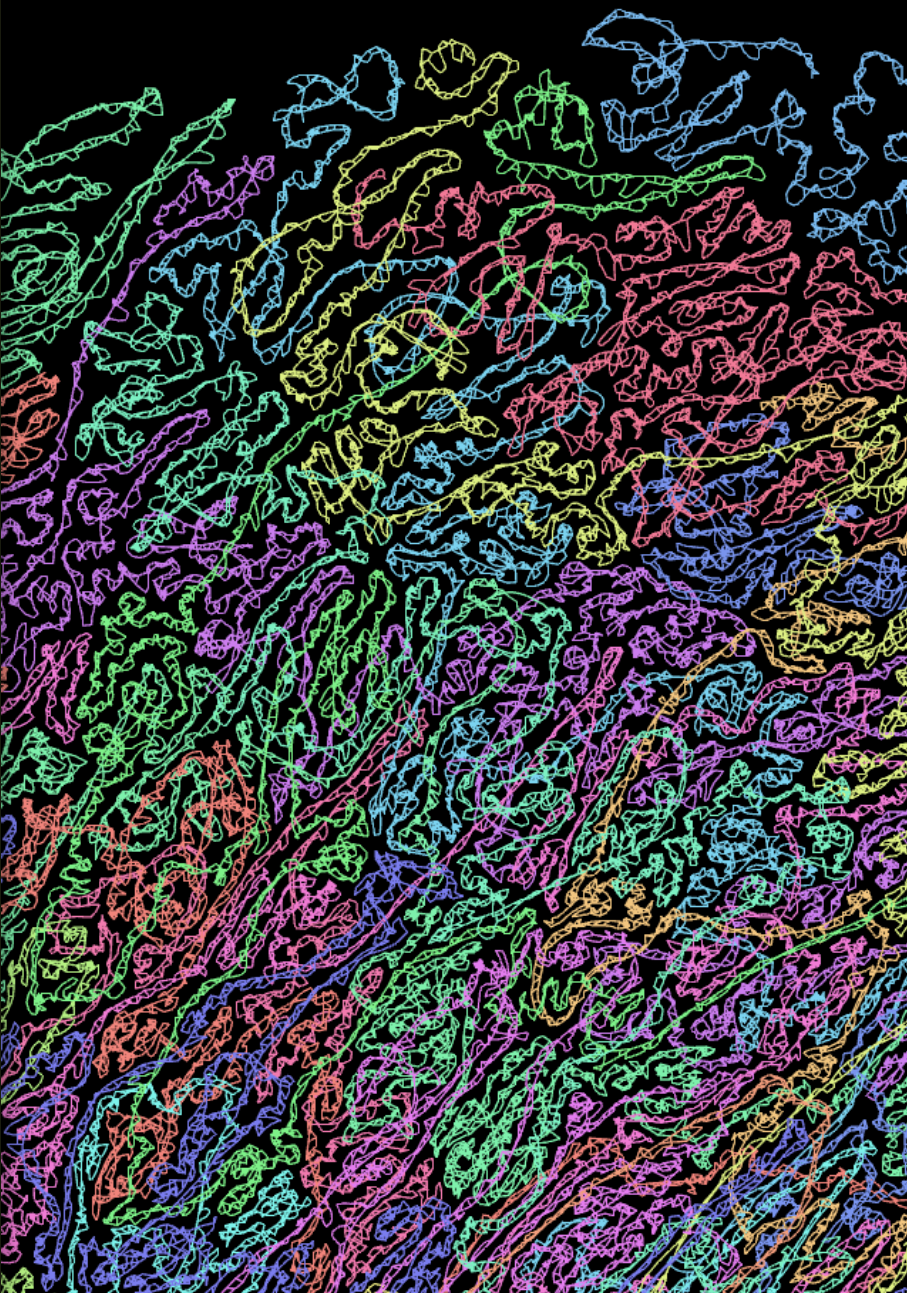
graph: Wheat - 1A exons: 22794 links: 25696



Labdesk Tools Library

local coarsen	local color-generator	local detectClusters	local lf
local ngraph.louvain	local onExpTSV	local onGFF	local renderF
local jstat	local spearson	local AutoExecuteLD	local onExonGFF





```
1  /**
2  * TODO:
3  * Transcript overlay for graphics
4  */
5
6  // on('bob', AutoExecuteLD);
7  // fire('bob', {value: 100}, 1)
8
9  const length = graph.getNodesCount();
10 let idx = [];
11 let bigData = {};
12 let counter = 0;
13 graph.on('kezdjed', handler)
14
15 graph.forEachNode((node) => {
16   var d = graph.getData(node.id);
17   if (d.expression) {
18     const gid = node.id;
19     delete d.expression.normalized;
20     for (const tr in d.expression.vectors) {
21       trid = gid + '.' + tr.split('_')[1];
22       bigData[trid] = d.expression.vectors[tr];
23       idx.push(trid);
24       if (node.id == 'TraesCS7B02G222100') {
25         console.log(bigData[trid])
26       }
27     }
28   }
29   counter++;
30   if (counter == length) graph.fire('kezdjed', idx)
31 }
32
33 function handler(idx) {
34   console.log(graph.getData('TraesCS7B02G222100'));
35   let ctx = 'const range = ' + JSON.stringify({from: 0, to: 10});
36   ctx += ',idx = ' + JSON.stringify(idx);
37   ctx += ',data = ' + JSON.stringify(bigData)
38   // console.log(ctx)
39   runFarm('callback(null, Object.keys(data).length)', ctx, 5, (err, out) => {
40     console.log('kakamano' + out)
41   })
42 }
43
44 // graph.forEachNode((n) => {
45 //   n.data = {color: '0x005dbb'};
46 //   if (graph.getData(n.id).strand == '-') n.data.color = '0x9055FF';
47 // })
```