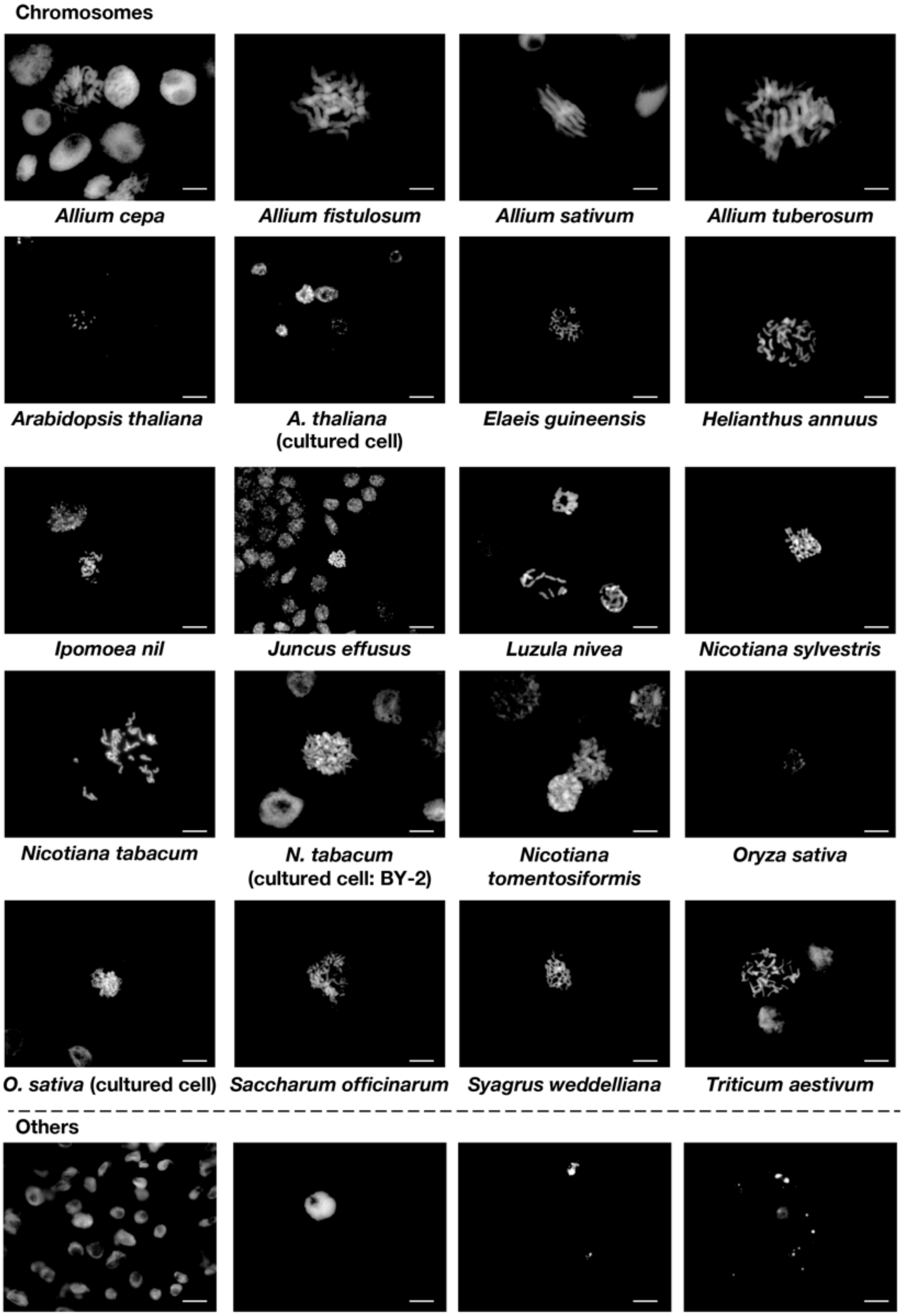
Title: Effectiveness of Create ML in microscopy image classifications: A simple and inexpensive deep learning pipeline for non-data scientists

Short title: A simple and inexpensive deep learning pipeline for non-data scientists



Chromosomes

Allium cepa | *Allium fistulosum* | *Allium sativum* | *Allium tuberosum*

Arabidopsis thaliana | *A. thaliana* (cultured cell) | *Elaeis guineensis* | *Helianthus annuus*

Ipomoea nil | *Juncus effusus* | *Luzula nivea* | *Nicotiana sylvestris*

Nicotiana tabacum | *N. tabacum* (cultured cell: BY-2) | *Nicotiana tomentosiformis* | *Oryza sativa*

O. sativa (cultured cell) | *Saccharum officinarum* | *Syagrus weddelliana* | *Triticum aestivum*

Others

Fig S1. Examples of data for classification of images with/without mitotic cells. Images with mitotic cells were annotated as "Chromosomes", and images without mitotic cells were annotated as "Others". The scale bar is 10 μm, but the images in the training and validation data did not include the scale bar.

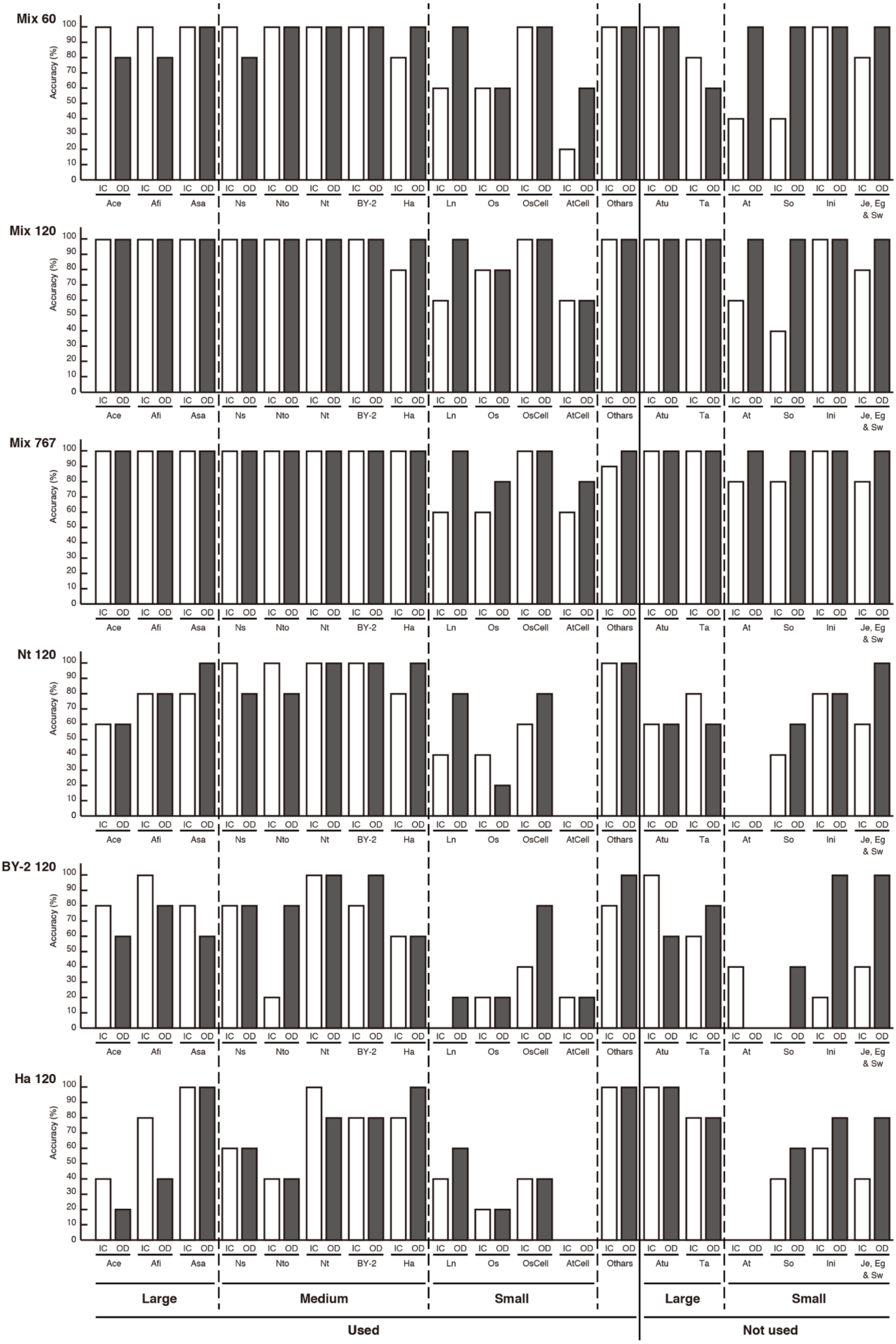
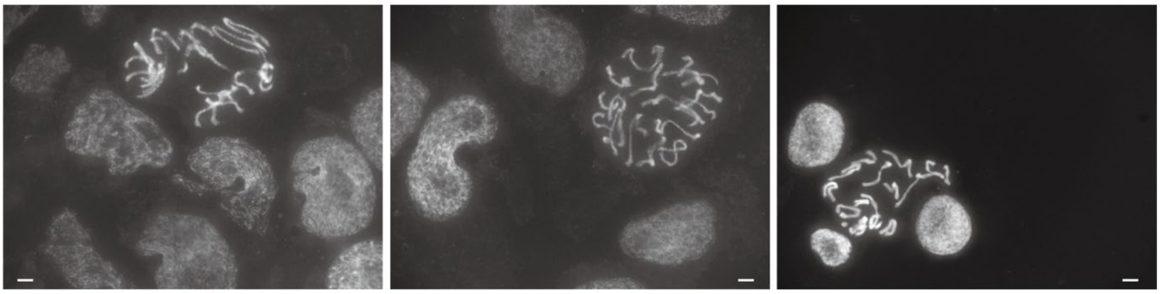

Mix 60
Mix 120
Mix 767
Nt 120
BY-2 120
Ha 120
Accuracy (%)
0
10
20
30
40
50
60
70
80
90
100
IC
OD
Ace
Afi
Asa
Ns
Nto
Nt
BY-2
Ha
Ln
Os
OsCell
AtCell
Othars
Atu
Ta
At
So
Ini
Je, Eg & Sw
Large
Medium
Small
Large
Small
Used
Not used

Fig S2. Details of the classification accuracy of the models constructed with different combinations of learning data. The methods of model construction and the types of learning data are shown below the graph. IC, image classifier without options and OD, object detector. “Large”, “medium” and “small” indicate the size of the chromosome of the species. "Used" and "not used" indicate whether the species was used for learning or not, respectively.

Chromosomes (Whole)



Chromosomes (Partial)

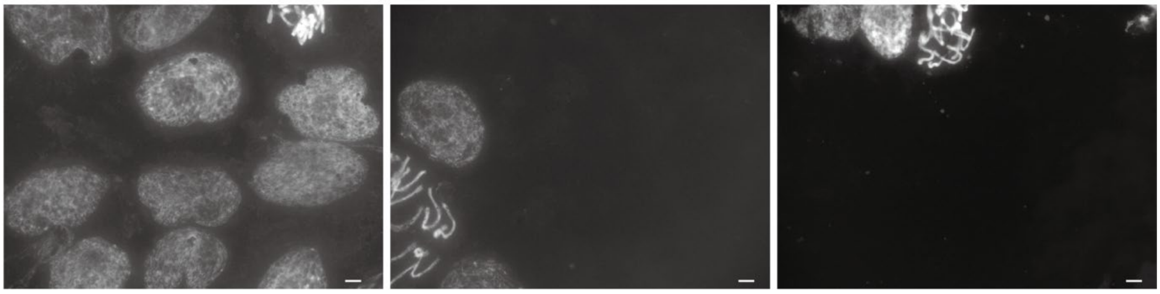


Fig S3. Examples of the validation data for classification of images with/without mitotic cells, captured by a Keyence microscope. Examples of images that include cells with whole chromosomes or partial chromosomes in the “Chromosomes” class. The scale bar is 10 μm, but the images in the training and validation data did not include the scale bar.

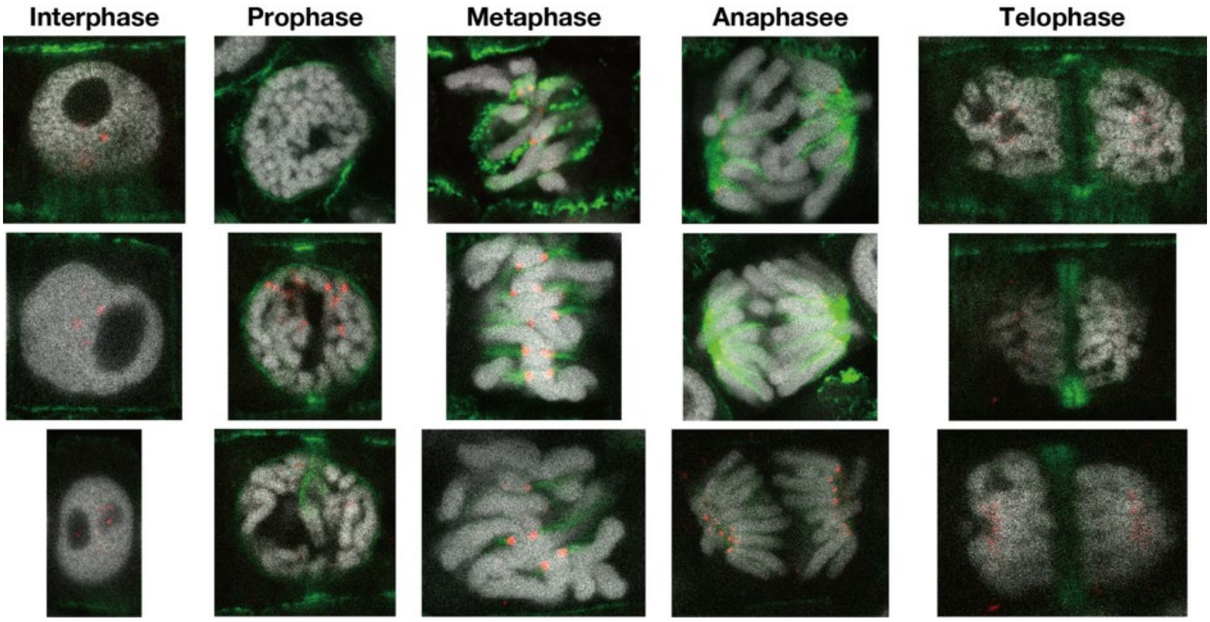


Fig S4. Examples of learning data for image classification of phases of the cell cycle.

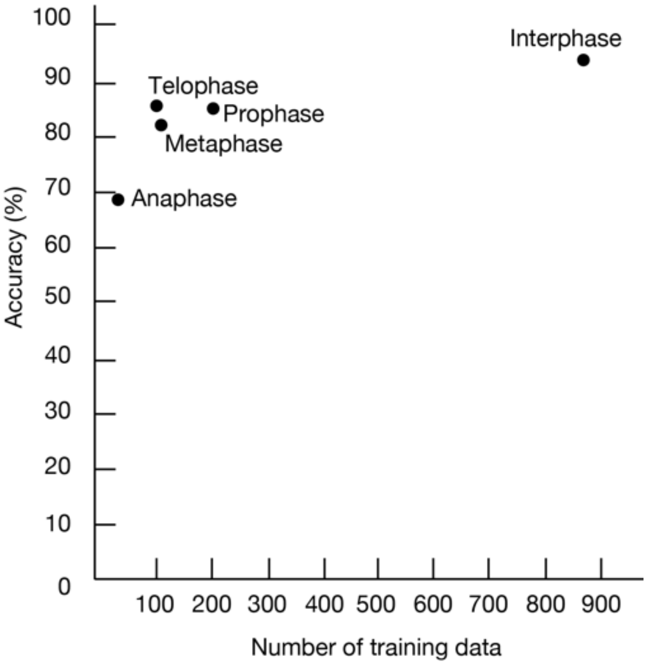


Fig S5. Relationship between the amount of learning data and the accuracy of image classification of the phases of the cell cycle.

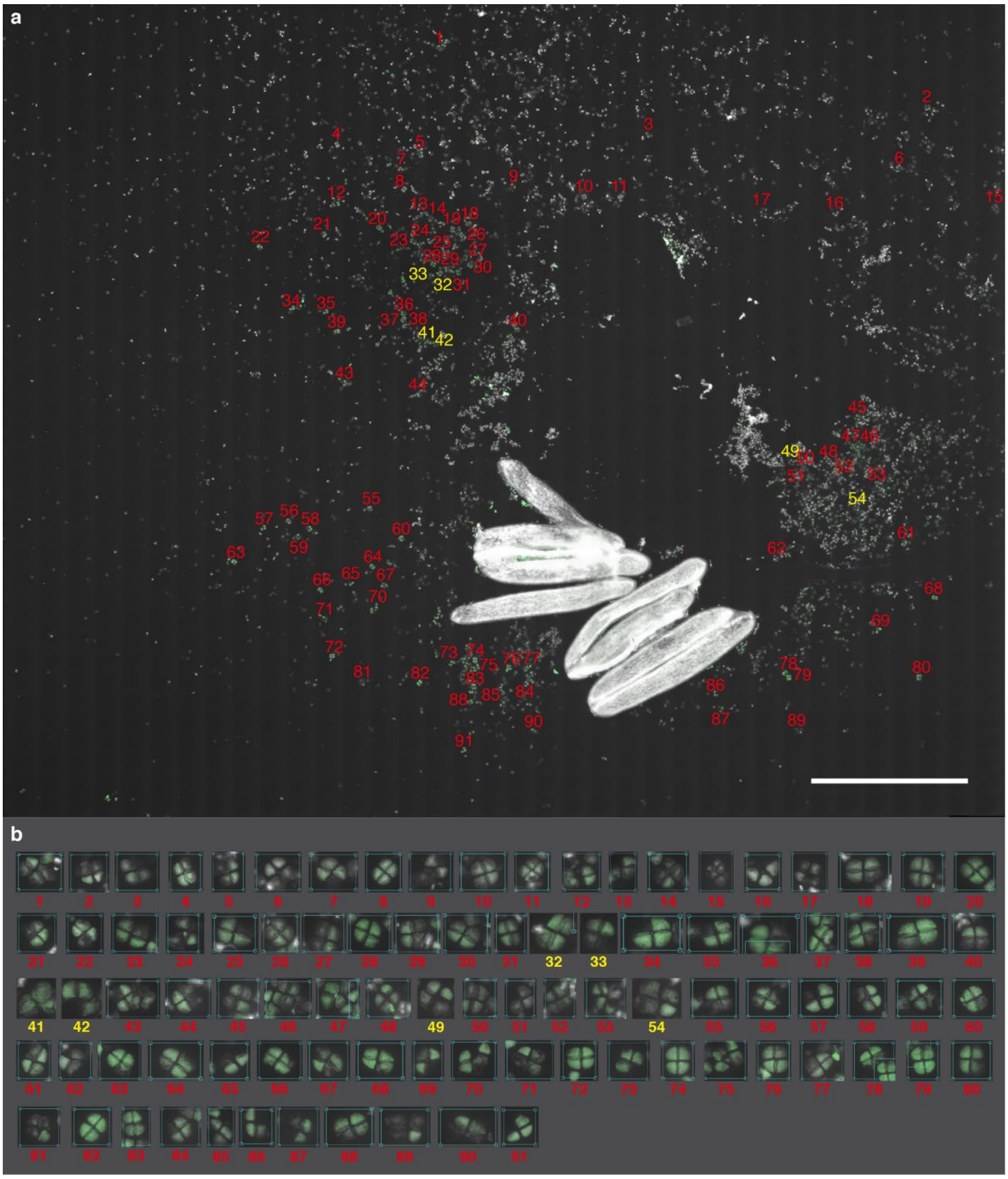


Fig S6. Identification of tetrads in a set of microscopy images. (a) Combined image of 1,404 (36 x 39 images) microscopy images. The positions of tetrads are indicated by numbers. The red and yellow numbers indicate tetrads that were or were not detected, respectively, by AI. The scale bar is 1 mm. (b) Images of the tetrads in the area. The areas annotated by AI are indicated by blue squares in the images. Note: Those numbers were added manually.

Table S1. Type and number of images used for the learning and validation datasets.

		Learning dataset												Validation dataset	
		Mix-60		Mix-120		Mix-767		Tobacco-120		BY-2-120		Sunflower-120			
Species/ cell name	Abbreviation	Chromosomes	Others	Chromosomes	Others	Chromosomes	Others	Chromosomes	Others	Chromosomes	Others	Chromosomes	Others	Chromosomes	Others
Allium cepa	Ace	5		10		29								5	
Allium fistulosum	Afi	5	15	10	30	31	59		30		30		30	5	3
Allium sativum	Asa	5		10		20								5	
Arabidopsis thaliana (cultured cell)	AtCell	5	15	10	30	13	61		30		30		30	5	2
Helianthus annuus	Ha	5	15	10	30	198	56		30		30	120	30	5	3
Luzula nivea	Ln	5		10		14								5	
Nicotiana sylvestris	Ns	5		10		15								5	
Nicotiana tomentosiformis	Nto	5		10		41								5	
Nicotiana tabacum	Nt	5		10		189		120						5	
N. tabacum	BY-2	5		10		187				120				5	

(cultured cell: BY-2)															
Oryza sativa	Os	5	15	10	30	15	56		30		30		30	5	2
O. sativa (cultured cell)	OsCell	5		10		15								5	
Allium tuberosum	Atu													5	
A. thaliana	At													5	
Elaeis guineensis	Eg													3	
Ipomoea nil	Ini													5	
Juncus effusus	Je													1	
Saccharum officinarum	So													5	
Syagrus weddelliana	Sw													1	
Triticum aestivum	Ta													5	
Total number		60	60	120	120	767	232	120	120	120	120	120	120	90	10

Table S2. Learning time and iterations of the datasets.

	Mix-60			Mix-120			Mix-767			Tobacco-120			BY-2-120			Sunflower-120		
	IC	IC+op	OD	IC	IC+op	OD	IC	IC+op	OD	IC	IC+op	OD	IC	IC+op	OD	IC	IC+op	OD
Learning time	28 s	1 h 44 min	1 h 18 min	59 s	3 h 49 min	2 h 12 min	5 min 6 s	15 h 4 min	4 h 28 min	1 min	3 h 36 min	1 h 34 min	59 s	3 h 48 min	1 h 58 min	59 s	3 h 31 min	2 h 9 min
Iterations	25	25	5,000	25	25	8,000	25	25	13,000	25	25	7,000	25	25	7,000	25	25	7,000

Table S3. Genome size, chromosome number and average chromosome length of the species used in this study.

Species	Genome size (Mbp)	Chromosome number (2*n*)	Average chromosome length (Mbp)
Allium cepa	17,542	16	1,096
Allium fistulosum	11,642	16	728
Allium sativum	15,876	16	992
Allium tuberosum	31,448	16	1,966
Arabidopsis thaliana	157	10	16
Elaeis guineensis	2,117	32	66
Helianthus annuus	3,597	34	106
Ipomoea nil	1,176	30	39
Juncus effusus	294	40	7
Luzula nivea	882	12	74
Nicotiana tabacum	5,072	48	106
Nicotiana tomentosiformis	2,688	24	112
Nicotiana sylvestris	2,641	24	110
Oryza sativa	490	24	20
Saccharum officinarum	3,724	80	47
Syagrus weddelliana	Unknown	32	Unknown
Triticum aestivum	16,954	42	404

Genome sizes are based on data from the Plant DNA C-values database (release 7.1) (Pellicer, 2020).

Supplementary movie 1. Overview of the deep learning pipeline

Supplementary file 1. Step-by-step protocol of the deep learning pipeline