

# カンキツのゲノム解析

柑橘産業イノベーションセンター 柑橘生産部門

愛媛大学大学院農学研究科

羽生 剛

# はじめに

## 1. ゲノム解析とは

ゲノム解析について概説します。

## 2. ゲノム解析の方法

ゲノム解析の方法を簡単に説明します。

## 3. カンキツのゲノム解析

カンキツを使ったゲノム解析の結果を紹介します。

# ゲノムとは

ゲノムとは,

GENE + OME = GENOME  
(遺伝子) (全体) (ゲノム)

基本的には遺伝子全体というのがもともとの意味



遺伝子とは？

# 遺伝子とは

遺伝子は、遺伝（親から子へ伝わる）する因子

例)

- ・花の色
- ・果実の形
- ・病気への耐性
- etc....

} 遺伝形質

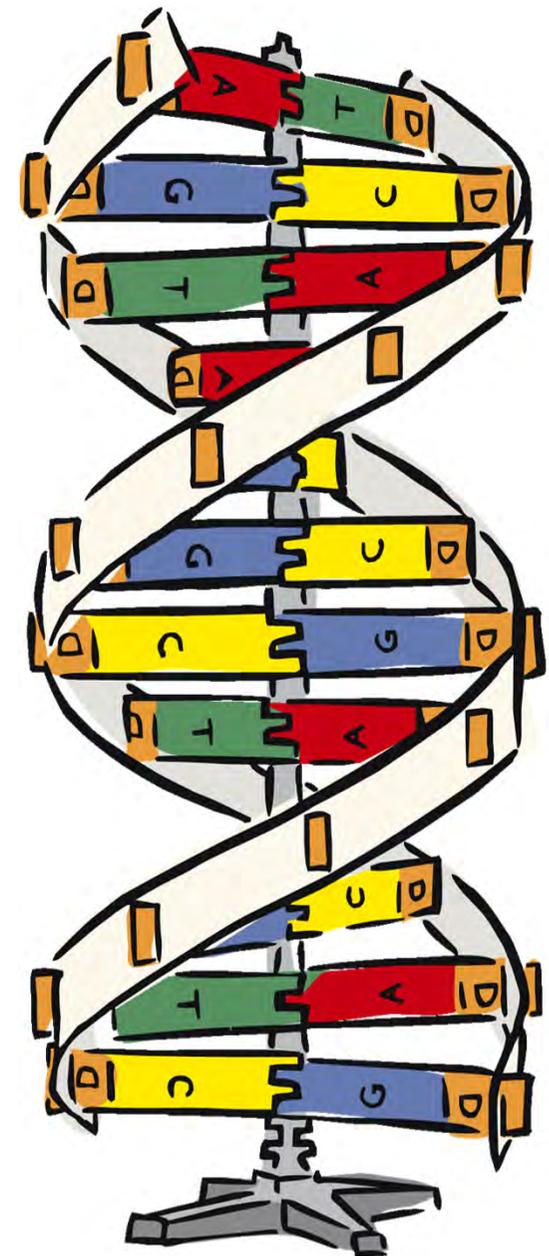
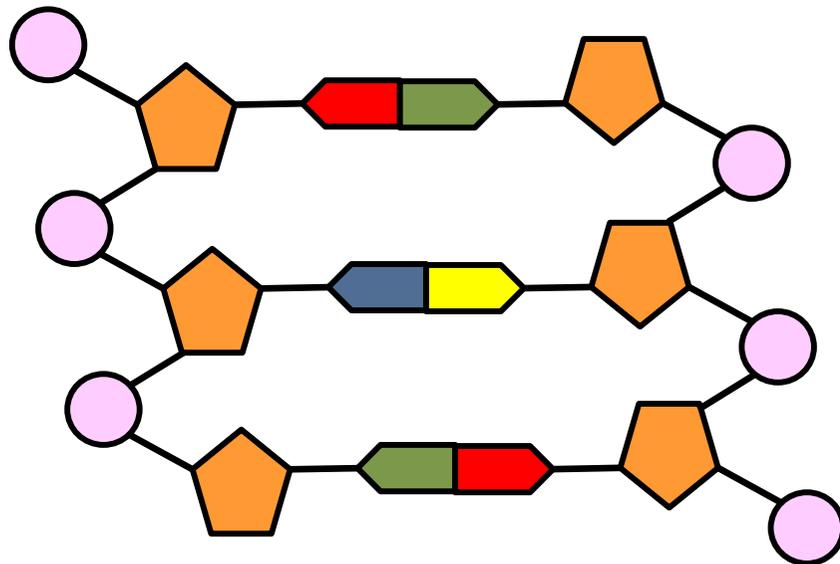
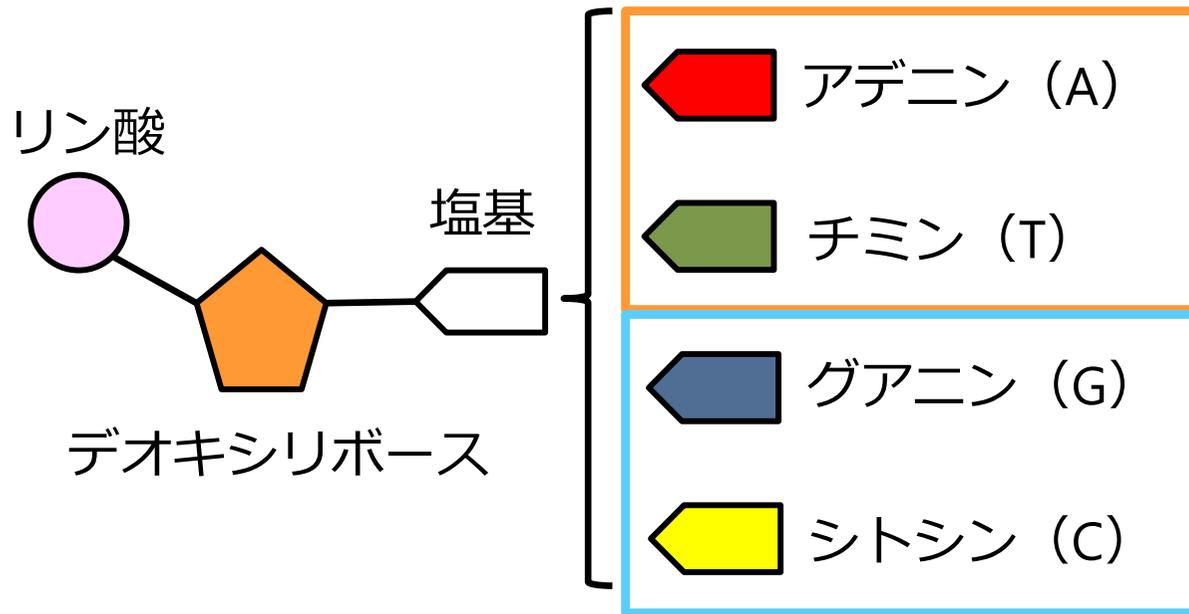


遺伝子は遺伝形質の情報が書かれた設計図

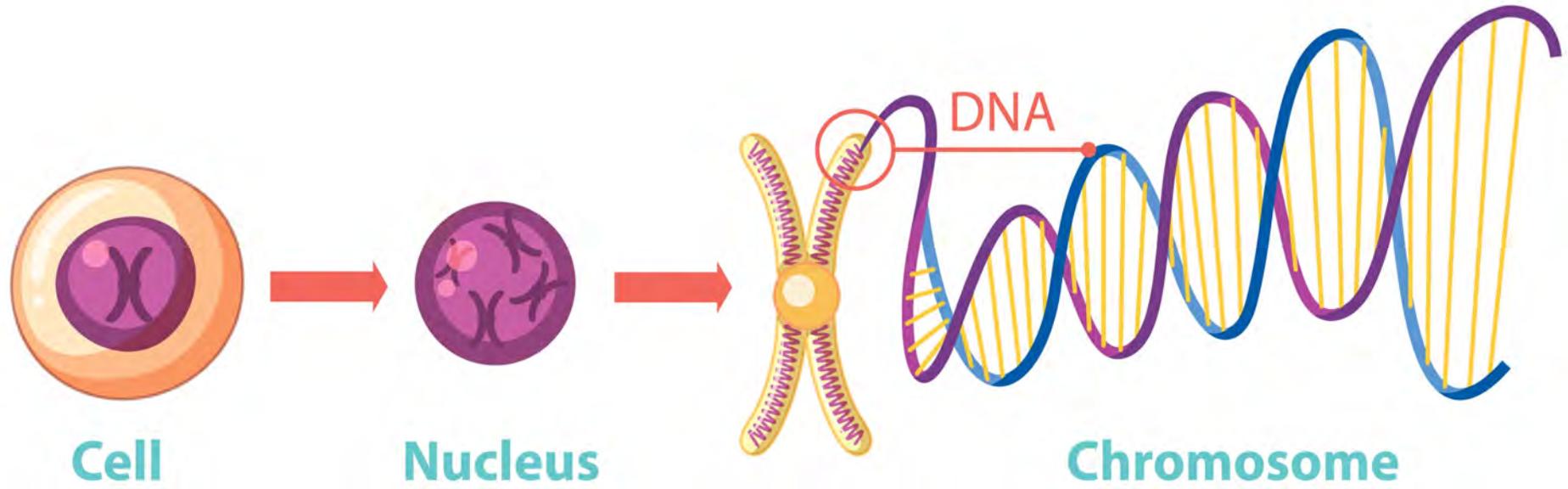


DNA（デオキシリボ核酸）

# DNAは4種類の塩基で構成されている



# DNAが折りたたまれたものが染色体



細胞

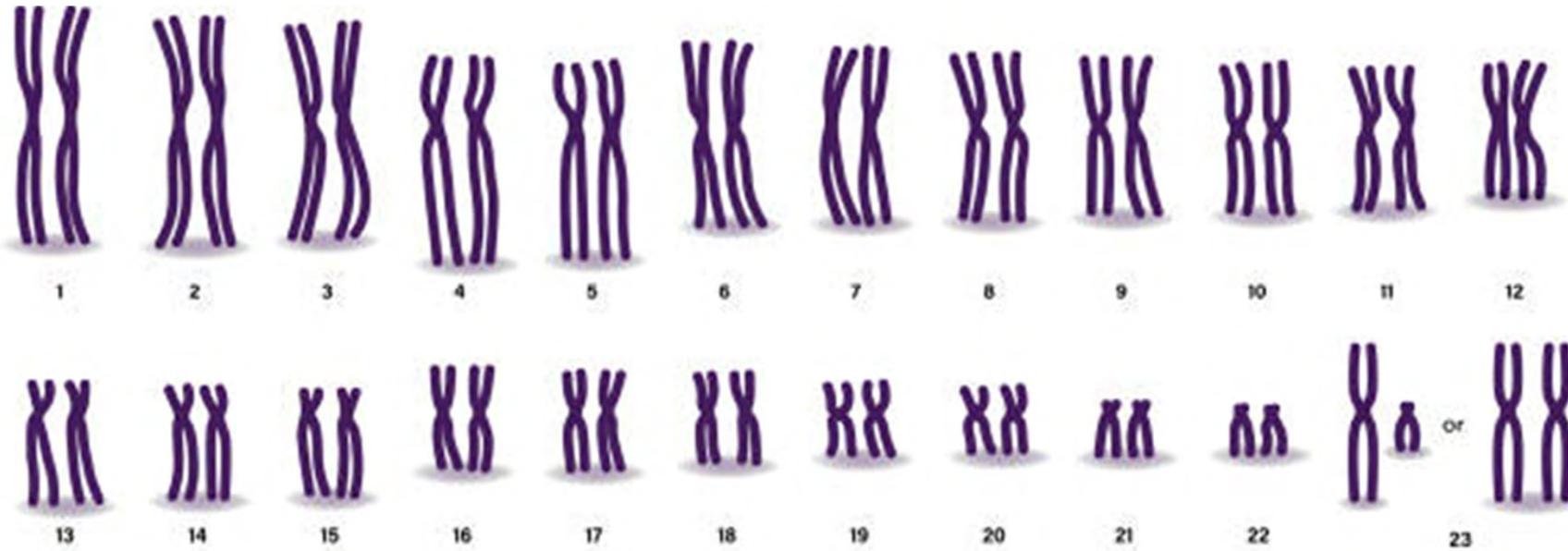
核

染色体

||

DNAがぐるぐる巻きになったもの

# 染色体 (=DNA) の1セットがゲノム



ヒトの場合、23対46本の染色体が各細胞の核内にある。

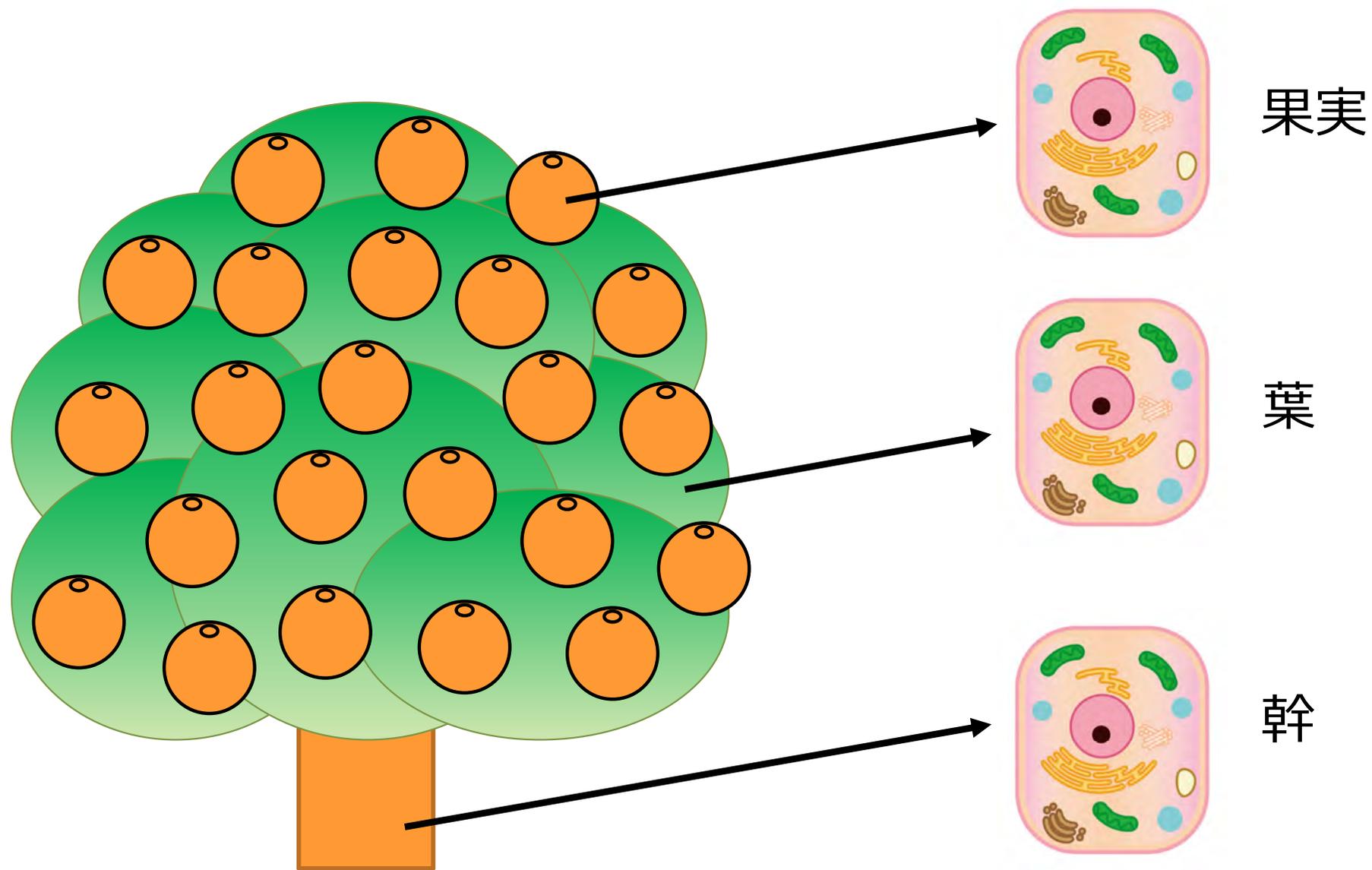


両親から23本で1セットを1セットずつ受け取り46本



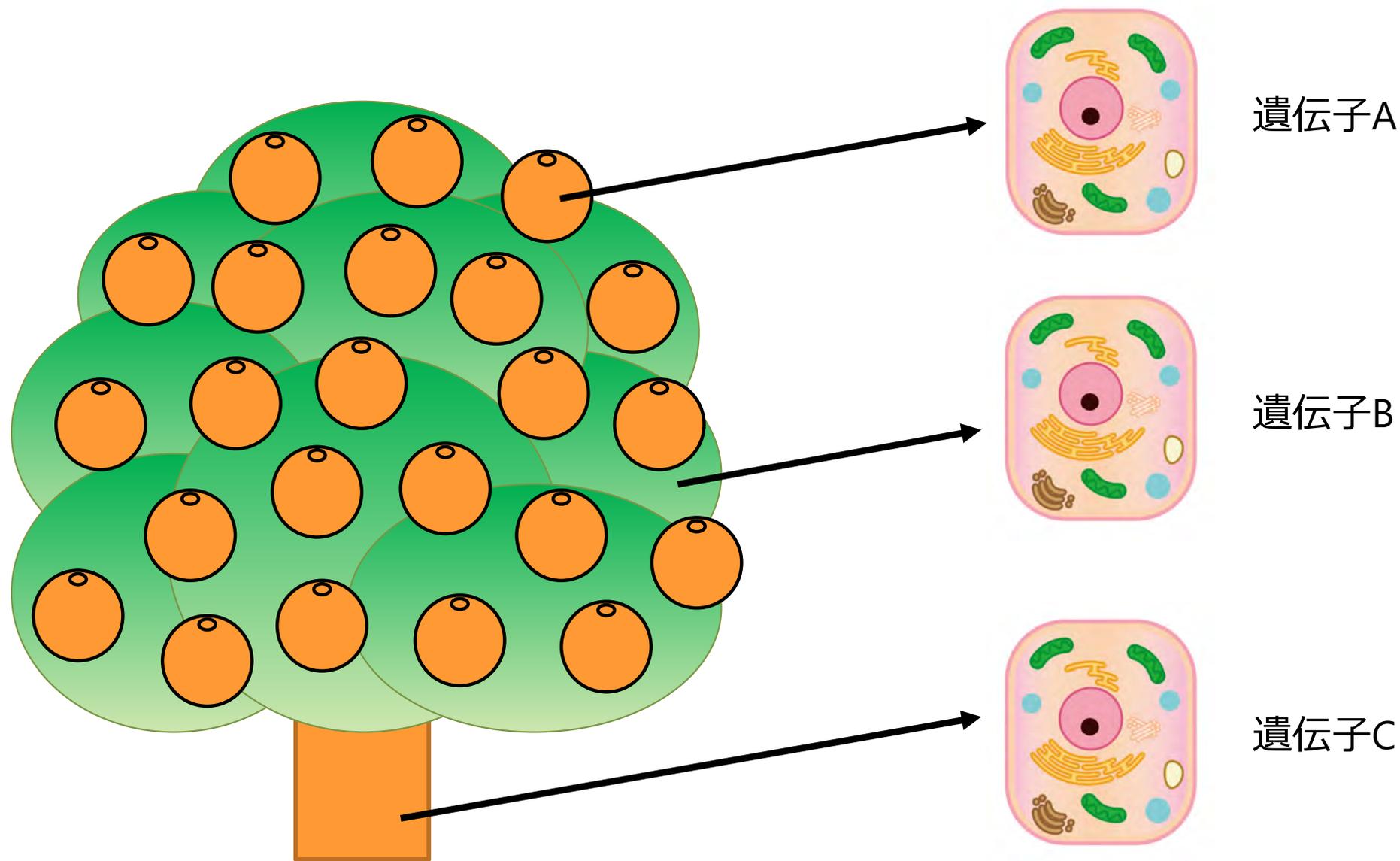
この1セット（23本の染色体 = 23個のDNA塩基配列）がゲノム  
（正確には、全染色体を構成するDNA配列の総体がゲノム）

# 全ての細胞が同じゲノムをもっている



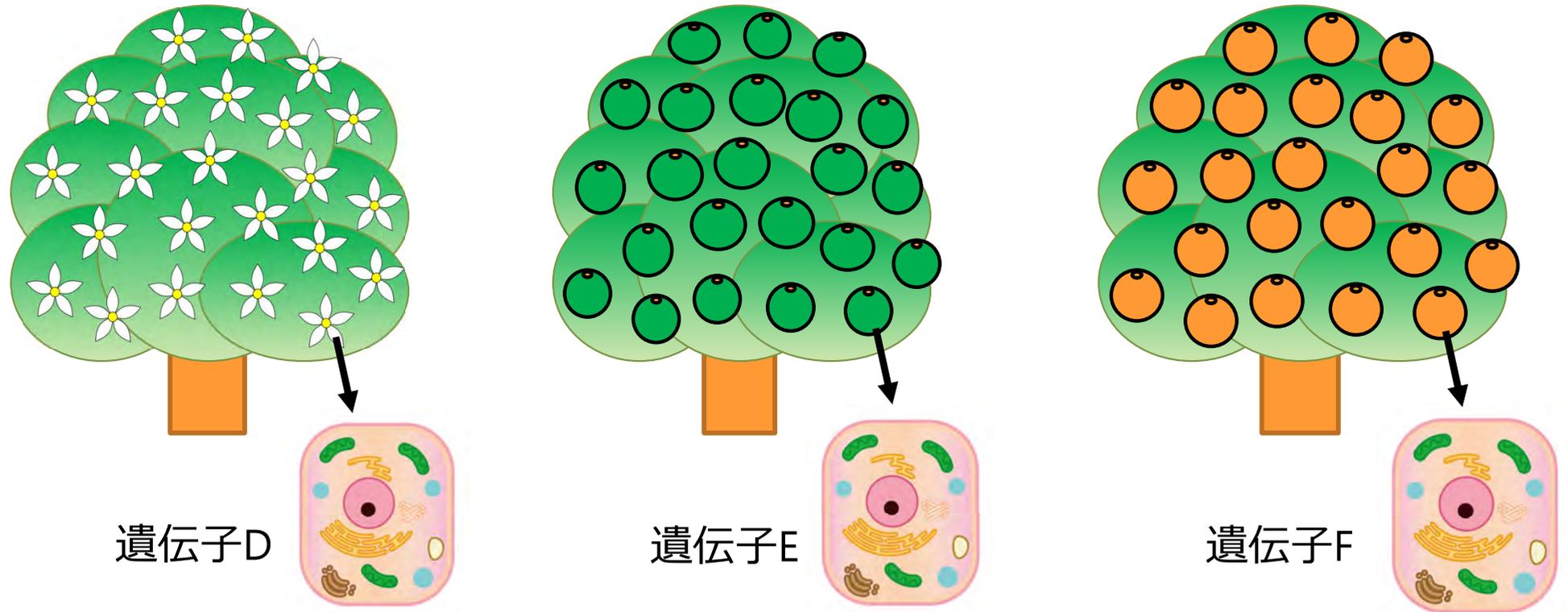
果実も葉も細胞に含まれる染色体（ゲノム）は同じ

# 遺伝子は働く場所が決まっている



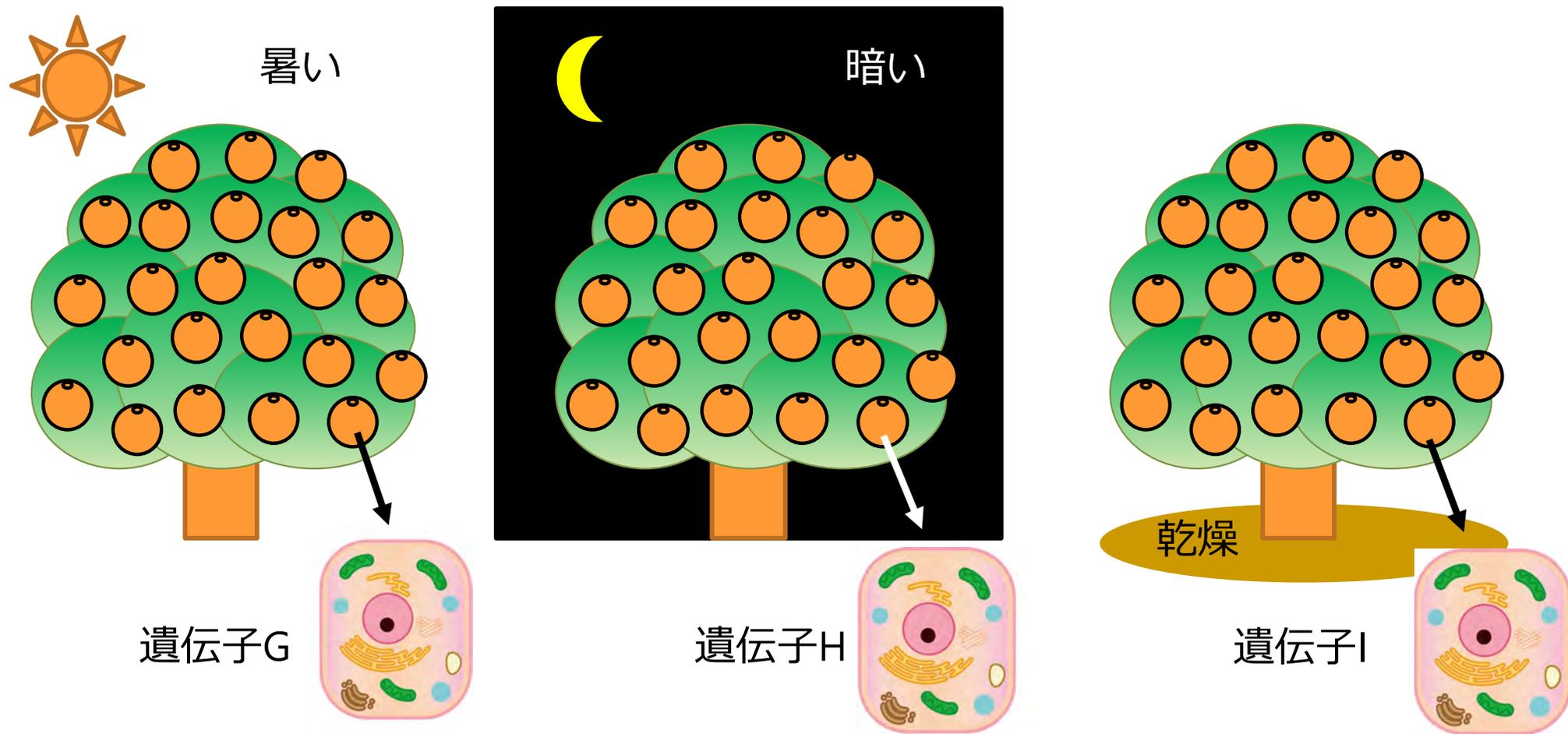
果実では果実の葉では葉の生育に必要な遺伝子が働く

# 遺伝子は働く時期が決まっている



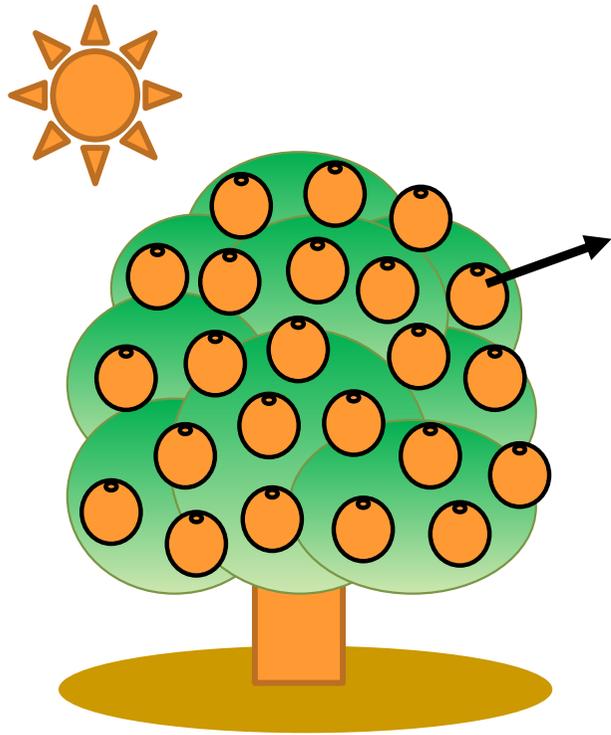
花では開花や生殖に必要な遺伝子が，幼果では果実の成長に必要な遺伝子が，成熟果では成熟に必要な遺伝子が働く。

# 遺伝子は環境の変化に応じて働く



暑いときには暑さに対応する遺伝子が、暗いときには暗いとき用の遺伝子が、土壌が乾燥している時には乾燥に対応する遺伝子が働く。

# 全ての遺伝子の働きがあわさったものが目に見える形で現れる



これまでの生育期間全体および現在の遺伝子の働きの全ての結果が目に見える形に現れている。

**遺伝子A**

遺伝子B

遺伝子C

遺伝子D

遺伝子E

**遺伝子F**

**遺伝子G**

遺伝子H

**遺伝子I**

樹齢

季節

生育ステージ

周囲の環境

- ・ 気温
- ・ 日照
- ・ 土壌水分
- など

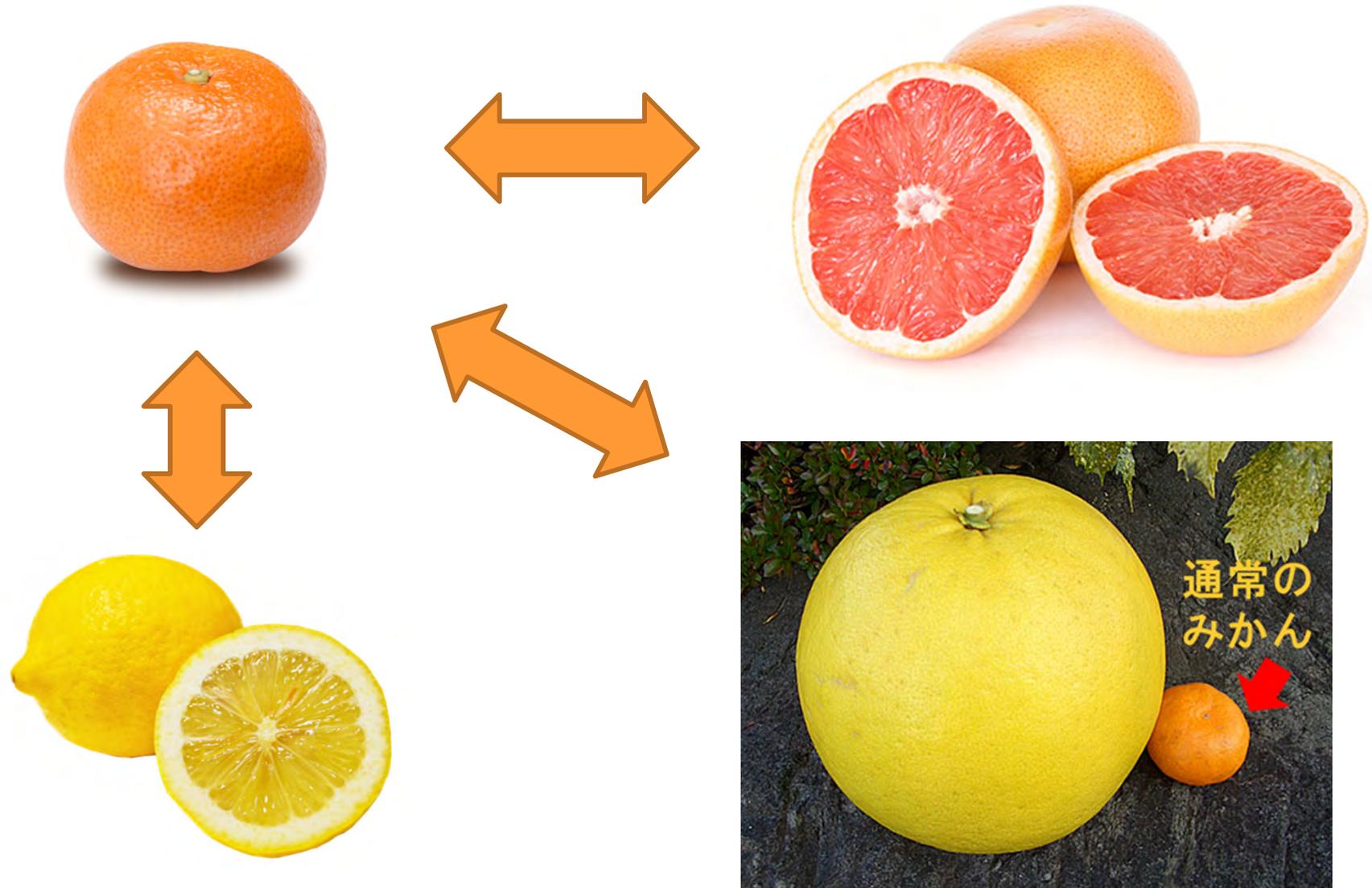
病虫害の被害

器官

他器官の状況

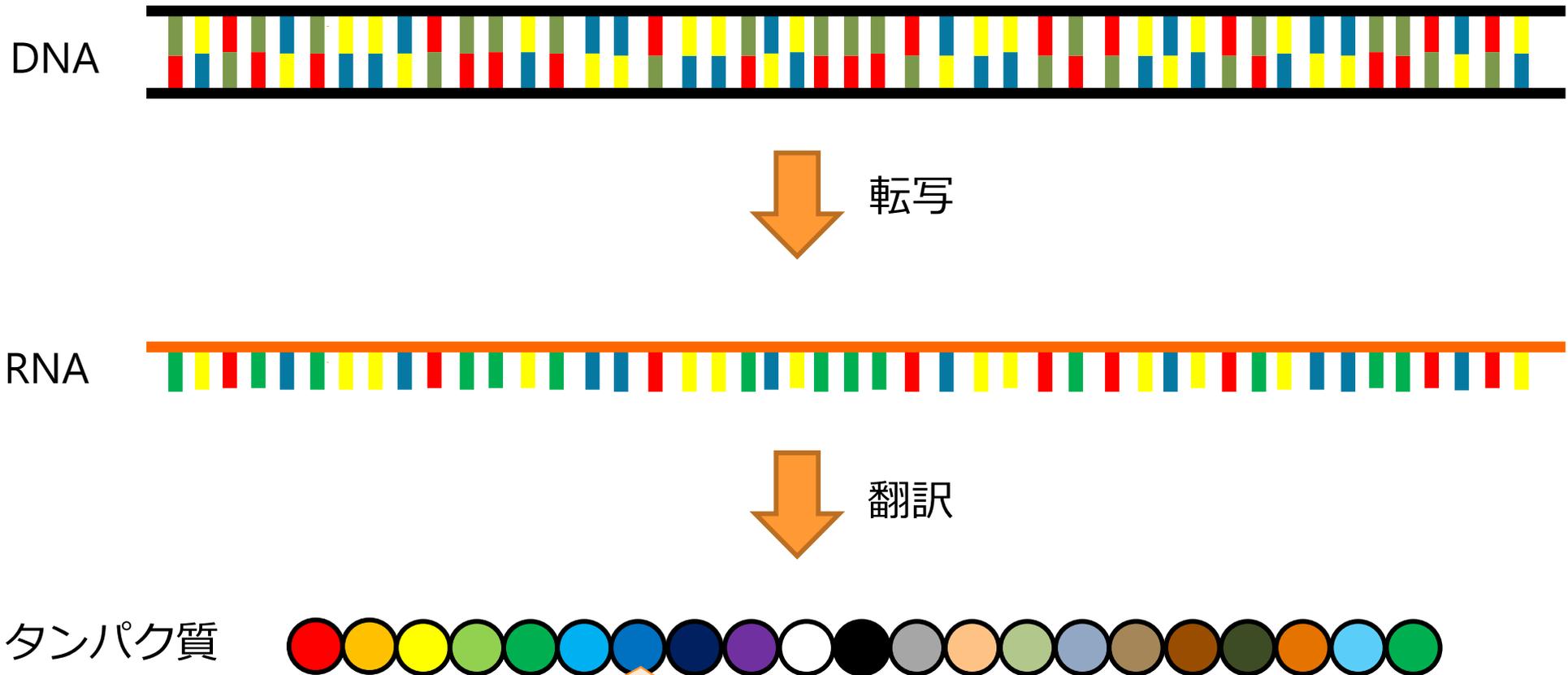
栄養成分

# 目に見える違いは遺伝子の働きの違い



様々な目に見える（測定可能なものも含む）違いは遺伝子の働きの違い

# 遺伝子が働くとは

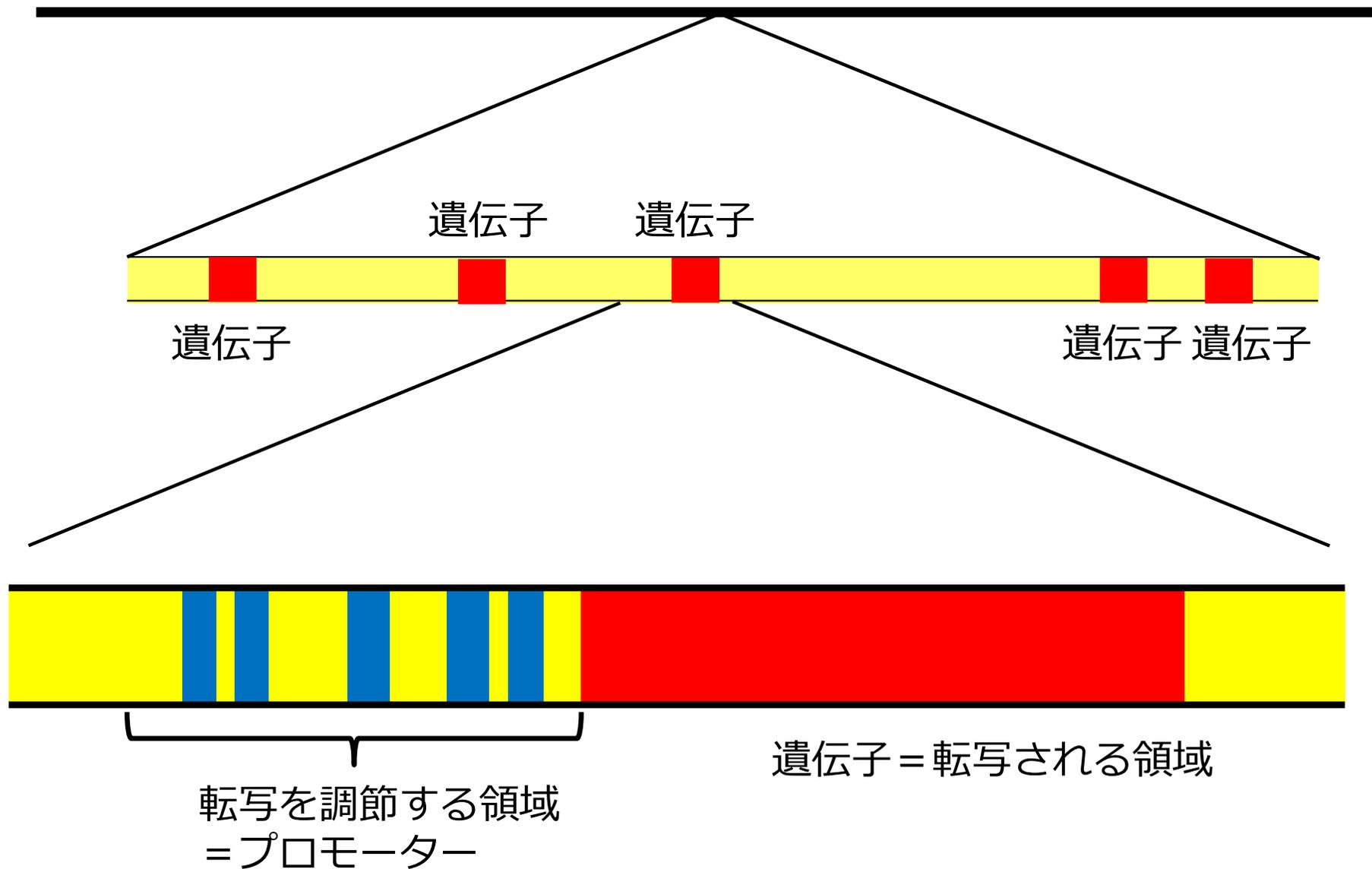


遺伝子が働く

||

タンパク質によって様々な生体内の化学反応が起こる

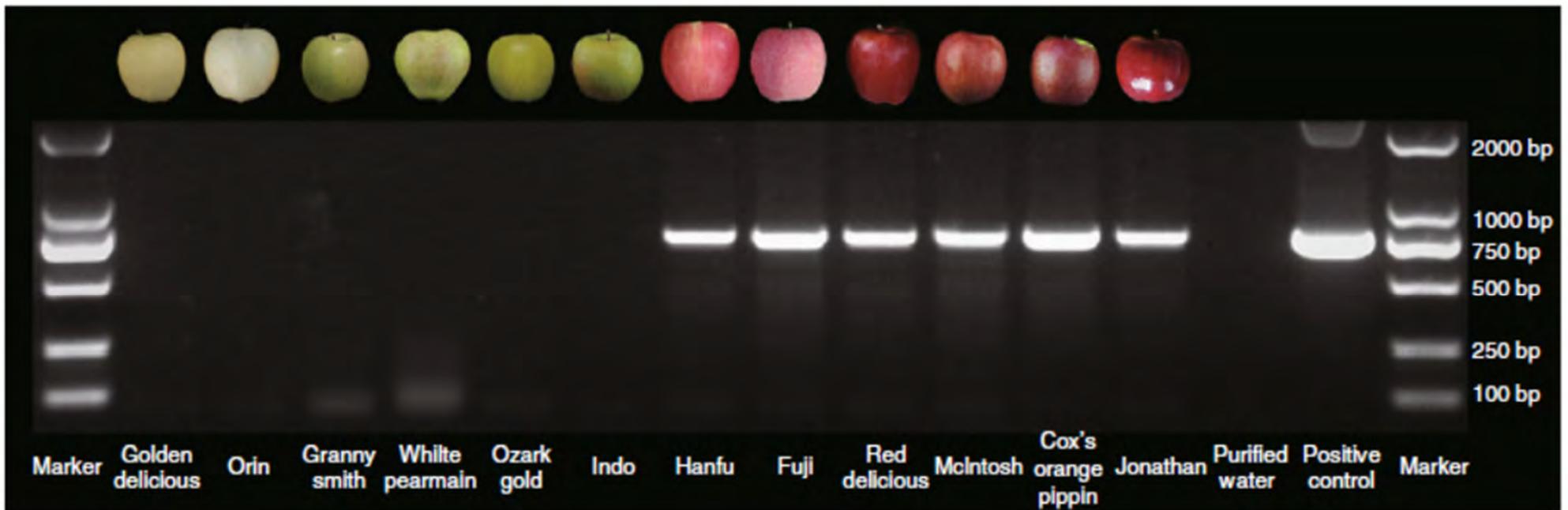
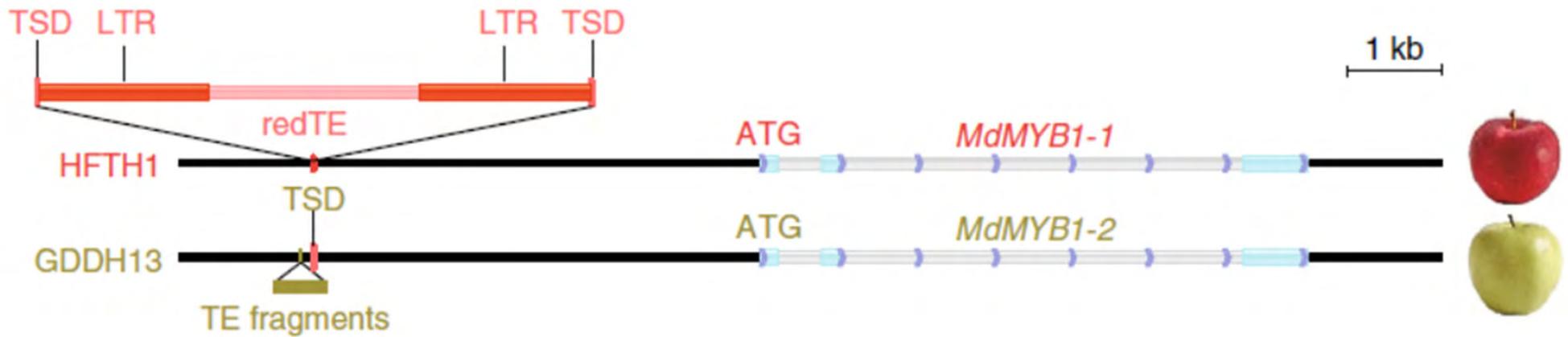
# 遺伝子の構造



# 遺伝子の働きの違いはDNAの塩基配列の違い

	転写	タンパク質	機能
 <p>塩基配列が欠損</p>	○	× (形が変わる)	×
	○	○	○
 <p>塩基配列が違う</p>	×	×	× (転写されない)

# 遺伝子の働きの違いはDNAの塩基配列の違い



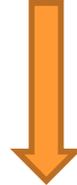
(Zhang et al., 2019)

# ゲノム解析は塩基配列を調べること

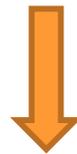
様々な現象面での違いはDNA配列の違いが原因



DNA配列（塩基配列）を調べることで様々な現象について明らかにできる



ゲノム解析



ゲノム解析でわかること

- ・ 果実形質に係する遺伝子がわかる
- ・ 病害虫や生理障害への耐性に関する遺伝子がわかる
- ・ 進化系統や親子関係がわかる

etc....

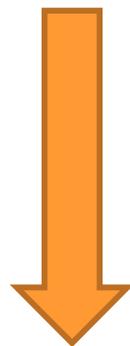
# ゲノム解析の方法

ゲノム解析は塩基配列を調べること



塩基配列を読む = シークエンス

しかし、ゲノムは大体サイズが大きい



例)

- ・ヒト 32億塩基
- ・マウス 28億塩基
- ・カンキツ 3億6千万塩基

長い配列を短い配列に分割して読む

# 次世代シーケンサー（第2世代）



Illumina Miseq

Illumina HiSeq

	Miseq	HiSeq
最大データ量	15Gb	1.5Tb (HiSeq4000) 1Tb (HiSeq2500)
最大リード数/ラン	5000万	100億 (HiSeq4000) 80億 (HiSeq2500)
最大リード長	300 b p x 2	150bp x 2 (HiSeq4000) 125bp x 2 (HiSeq2500)

読める配列は短いですが、一度に大量の配列を並行して読めることが特徴

# シークエンスの結果

## 紅まどんな

AAAAAAAAAAAGGCGG  
CCCCCAAAAAAAAAA  
TTCCCCCCCCCAA  
GGGGTTTTTTTTTTC  
TTTTCCCCCCCCCA  
AAGGCGGGGGGGT  
CCCCCCCCAAAAAA  
GGGGGGGTTTTTTT  
AAAAGGCGGGGGGG  
GTTTTTTTTTTCCC  
⋮

## 甘平

CCCCCCCCAAAAAA  
GGGGTATTTTTTTC  
AAAAGGGGGGGGGG  
TTTTCCCCCCCCCA  
GTATTTTTTTTTCCC  
TTCCCCCCCCCAA  
CCCCCAAAAAAAAAA  
AAGGGGGGGGGGT  
AAAAAAAAAAGGGGG  
GGGGGGGTATTTTT  
⋮

# カンキツのゲノム解読

2000年 シロイヌナズナ解読完了

2003年 ヒトゲノム解読完了

2004年 イネ解読完了

- 
- 
- 

2013年 オレンジゲノム解読完了

2014年 クレメンティンゲノム解読完了

2017年 ブンタンゲノム解読完了  
ウンシュウミカンゲノム解読完了

ARTICLES

nature genetics  
OPEN

## The draft genome of sweet orange (*Citrus sinensis*)

Qiang Xu<sup>1,6</sup>, Ling-Ling Chen<sup>2,6</sup>, Xiaolan Ruan<sup>3,6</sup>, Dijun Chen<sup>2</sup>, Andan Zhu<sup>1</sup>, Chunli Chen<sup>1,2</sup>, Denis Bertrand<sup>3</sup>, Wen-Biao Jiao<sup>2</sup>, Bao-Hai Hao<sup>2</sup>, Matthew P Lyon<sup>4</sup>, Jiongjiong Chen<sup>1</sup>, Song Gao<sup>1</sup>, Feng Xing<sup>2</sup>, Hong Lan<sup>1</sup>, Ji-Wei Chang<sup>2</sup>, Xianhong Ge<sup>5</sup>, Yang Lei<sup>2</sup>, Qun Hu<sup>1</sup>, Yin Miao<sup>2</sup>, Lun Wang<sup>1</sup>, Shixin Xiao<sup>4</sup>, Manosh Kumar Biswas<sup>1</sup>, Wenfang Zeng<sup>1</sup>, Fei Guo<sup>1</sup>, Hongbo Cao<sup>1</sup>, Xiaoming Yang<sup>1</sup>, Xi-Wen Xu<sup>2</sup>, Yun-Jiang Cheng<sup>1</sup>, Juan Xu<sup>1</sup>, Ji-Hong Liu<sup>1</sup>, Oscar Junhong Luo<sup>3</sup>, Zhonghui Tang<sup>2</sup>, Wen-Wu Guo<sup>1</sup>, Hanhui Kuang<sup>1</sup>, Hong-Yu Zhang<sup>2</sup>, Mikeal L. Roose<sup>4</sup>, Niranjan Nagarajan<sup>3</sup>, Xiu-Xin Deng<sup>1</sup> & Yijun Ruan<sup>2,3</sup>

Oranges are an important nutritional source for human health and have immense economic value. Here we present a comprehensive analysis of the draft genome of sweet orange (*Citrus sinensis*). The assembled sequence covers 87.3% of the estimated orange genome, which is relatively compact, as 20% is composed of repetitive elements. We predicted 29,445 protein-coding genes, half of which are in the heterozygous state. With additional sequencing of two more citrus species and comparative analyses of seven citrus genomes, we present evidence to suggest that sweet orange originated from a backcross hybrid between pummelo and mandarin. Focused analysis on genes involved in vitamin C metabolism showed that *GaLUR*, encoding the rate-limiting enzyme of the galacturonate pathway, is significantly upregulated in orange fruit, and the recent expansion of this gene family may provide a genomic basis. This draft genome represents a valuable resource for understanding and improving many important citrus traits in the future.

*Citrus* is a large genus that includes several major cultivated species, including *C. sinensis* (sweet orange), *Citrus reticulata* (tangerine and mandarin), *Citrus limon* (lemon), *Citrus grandis* (pummelo) and *Citrus paradisi* (grapefruit). In 2009, the global citrus acreage was 9 million hectares and citrus production was 122.3 million tons (FAO statistics, see URLs), which is the top ranked among all the fruit crops. Among the 10.9 million tons (valued at \$9.3 billion) of citrus products traded in 2009, sweet orange accounted for approximately 60% of citrus production for both fresh fruit and processed juice consumption (FAO statistics, see URLs). Moreover, citrus fruits and juice are the prime human source of vitamin C, an important component of human nutrition.

Citrus fruits also have some unique botanical features, such as nucellar embryony (nucellus cells can develop into apomictic embryos that are genetically identical to mother plant). Consequently, somatic embryos grow much more vigorously than the zygotic embryos in seeds such that seedlings are essentially clones of the maternal parent. Such citrus-unique characteristics have hindered the study of citrus genetics and breeding improvement<sup>1,2</sup>. Complete genome sequences would provide valuable genetic resources for improving citrus crops.

*Citrus* is believed to be native to southeast Asia<sup>3-5</sup>, and cultivation of fruit crops occurred at least 4,000 years ago<sup>3,6</sup>. The genetic origin of the sweet orange is not clear, although there are some speculations that sweet orange might be derived from interspecific hybridization of some primitive citrus species<sup>7,8</sup>. *Citrus* is also in the order Sapindales, a sister order to the Brassicales in the Malvales, making it valuable for comparative genomics studies with the model plant *Arabidopsis*.

We aimed to sequence the genome of Valencia sweet orange (*C. sinensis* cv. Valencia), one of the most important sweet orange varieties cultivated worldwide and grown primarily for orange juice production. Normal sweet oranges are diploids, with nine pairs of chromosomes and an estimated genome size of ~367 Mb<sup>9</sup>. To reduce the complexity of the sequenced genome, we obtained a double-haploid (dihaploid) line derived from the anther culture of Valencia sweet orange<sup>10</sup>. We first generated whole-genome shotgun paired-end-tag sequence reads from the dihaploid genomic DNA and built a *de novo* assembly as the citrus reference genome; we then produced shotgun sequencing reads from the parental diploid DNA and mapped the sequences to the haploid reference genome to obtain the complete genome information for Valencia sweet orange. In addition, we conducted comprehensive transcriptome sequencing analyses for four representative tissues using shotgun RNA sequencing (RNA-Seq) to capture all transcribed sequences and paired-end-tag RNA sequencing (RNA-PET) to demarcate the 5' and 3' ends of all transcripts. On the basis of the DNA and RNA sequencing data, we characterized the orange genome for its gene content, heterozygosity and evolutionary features. The genome and transcriptome analyses presented in this study yield new insights into the origin of sweet orange and the genomic basis of vitamin C metabolism and provide a rich resource of genetic information for citrus breeding and genetic improvement.

<sup>1</sup>Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China. <sup>2</sup>College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China. <sup>3</sup>Genome Institute of Singapore, Singapore. <sup>4</sup>Botany and Plant Sciences, University of California, Riverside, California, USA. <sup>5</sup>College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China. <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to X.-X.D. (xxdang@mail.hzau.edu.cn) or Y.R. (yijun.ruan@gmail.com).

Received 30 April; accepted 24 October; published online 25 November 2012; doi:10.1038/ng.2472

# 参照配列との比較 (マッピング)

AAAAAAAAAAGGGGGGGGGGTTTTTTTTTTCCCCCCCCCAAAAAAAAAA  
AAAAAAAAAAGGCGG  
AAAAAGGCGGGGGGGG  
AAGGCGGGGGGGGTTT  
↑GGGGGGGGTTTTTTTTT  
GGGGTTTTTTTTTTTC  
GTTTTTTTTTTTCCC  
TTTTCCCCCCCCCA  
TTCCCCCCCCCAA  
CCCCCCCCAAAAAA  
CCCCCAAAAAAAAAA

紅まどんな

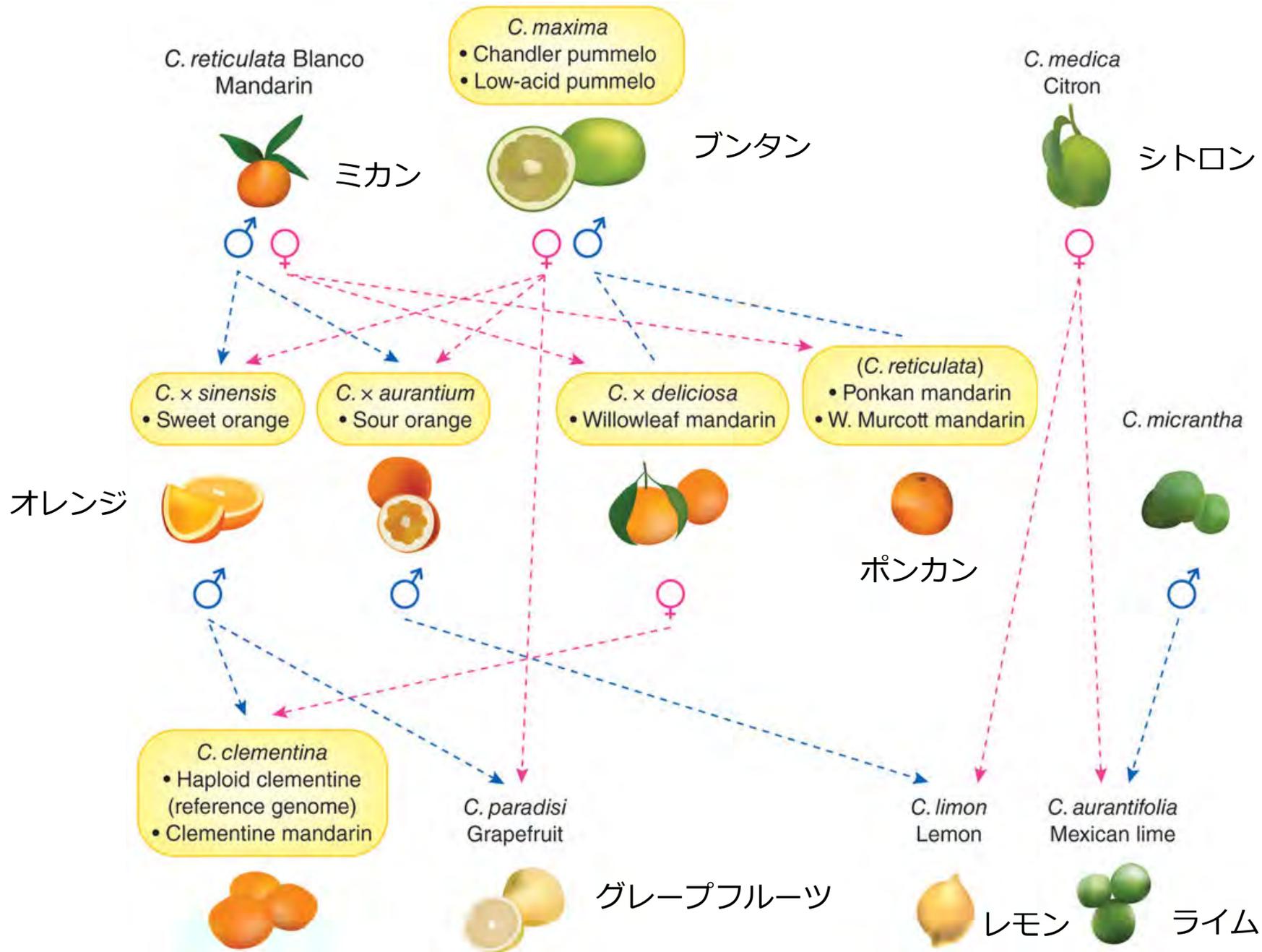
AAAAAAAAAAGGGGGG  
AAAAAGGGGGGGGGG  
AAGGGGGGGGGGTAT  
GGGGGGGGTATTTTTT  
GGGGTATTTTTTTTC  
GTATTTTTTTTCCC  
↑TTTTCCCCCCCCCA  
TTCCCCCCCCCAA  
CCCCCCCCAAAAAA  
CCCCCAAAAAAAAAA

甘平

参照配列に沿って並べると品種間の違いがわかる

# 中晩柑のゲノム解析

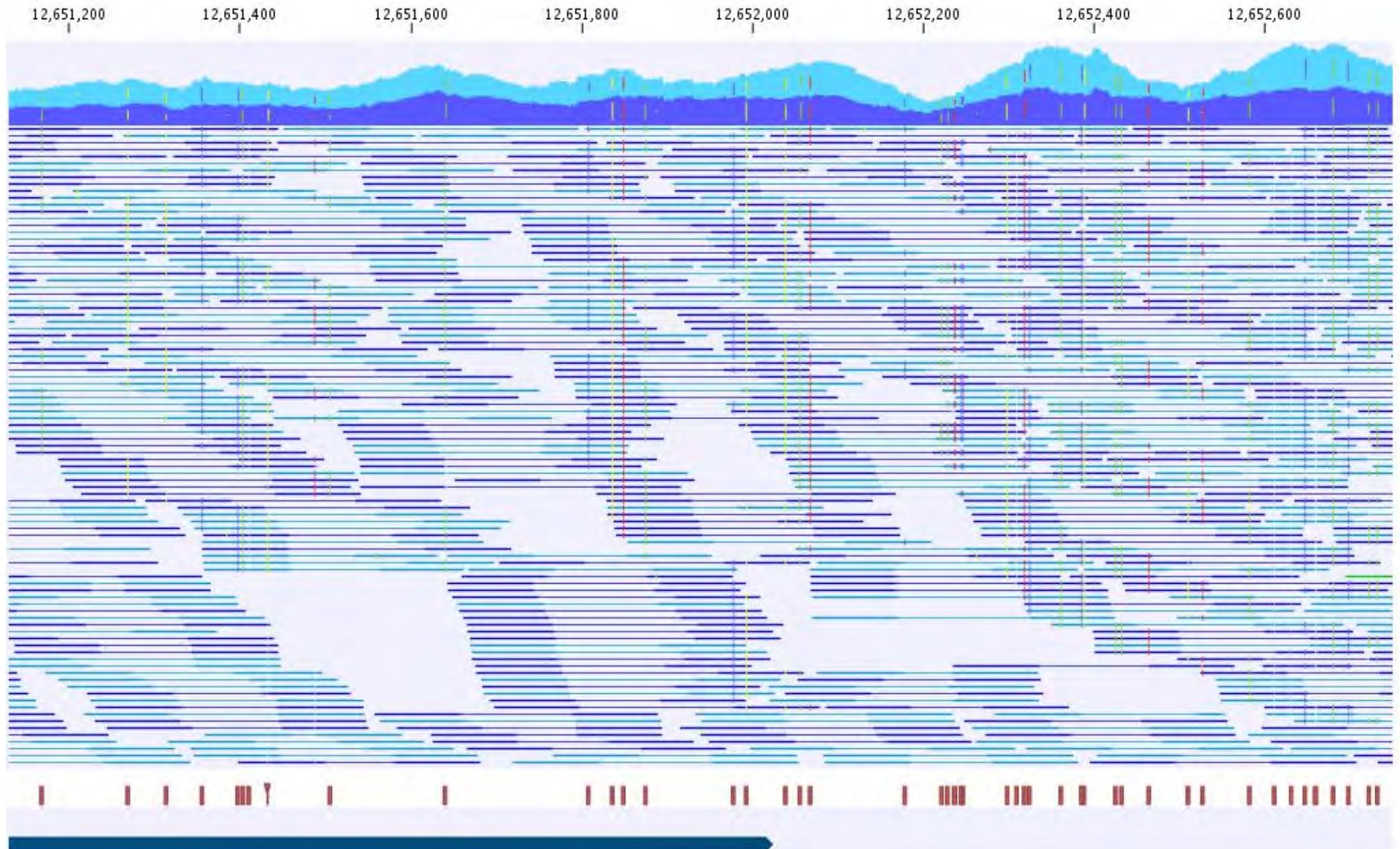
# カンキツの系統発生



# 供試材料

品種	種名	Acc. No.	トリム後リード数
ダンシー	<i>C. tangerina</i>	SRR5128277	70,745,932
無核紀州	<i>C. kinokuni</i>	SRR5128231	100,389,436
クレオパトラ	<i>C. reshni</i>	SRR5128221	142,819,786
ポンカン	<i>C. reticulata</i>	SRR1023619	182,120,812
クレメンティン	<i>C. clementina</i>	SRR1022654	416,675,884
マーコット	<i>C. reticulata</i>	SRR1023626	146,888,910
日南1号	<i>C. unshiu</i>	This study	78,338,336
宮川早生	<i>C. unshiu</i>	This study	86,092,020
興津早生	<i>C. unshiu</i>	This study	86,864,650
南柑20号	<i>C. unshiu</i>	This study	75,850,062
石地	<i>C. unshiu</i>	This study	79,359,484
青島	<i>C. unshiu</i>	This study	73,769,332
甘平	<i>C. spp.</i>	This study	82,767,012
愛媛果試第28号	<i>C. spp.</i>	This study	75,728,334
せとか	<i>C. spp.</i>	This study	77,467,570
清見	<i>C. spp.</i>	This study	90,146,940
スイートオレンジ	<i>C. sinensis</i>	SRR4240447	144,256,754
ワシントンネーブル	<i>C. sinensis</i>	SRR1023639	527,761,002
トロビタオレンジ	<i>C. sinensis</i>	DRP142816	185,797,460
ダンカン	<i>C. paradisi</i>	SRR128227	120,394,740
河内晩柑	<i>C. spp.</i>	This study	82,187,750
麻豆ブンタン	<i>C. grandis</i>	SRR5128244	67,354,480
Hunan pummelo	<i>C. grandis</i>	SRR5128240	92,713,392
Tahitian pummelo	<i>C. grandis</i>	SRR5128237	74,938,404
カオパン	<i>C. grandis</i>	SRR5128236	149,244,244

# マッピングしてみると



# マッピングの結果

品種	ウンシュウミカン		スイートオレンジ		ブンタン	
	全体	遺伝子領域	全体	遺伝子領域	全体	遺伝子領域
ダンシー	2,163,172	541,424	2,775,388	863,575	4,496,092	1,324,796
無核紀州	2,043,018	490,544	3,353,824	987,113	5,322,515	1,483,660
クレオパトラ	3,161,430	600,042	4,247,884	1,079,324	7,061,333	1,686,812
ポンカン	3,988,517	758,945	4,279,110	1,093,672	7,431,384	1,725,458
クレメンティン	4,589,570	888,356	3,735,408	1,024,533	7,374,654	1,705,257
マーコット	3,594,976	707,133	3,118,197	902,792	6,765,199	1,616,406
日南1号	2,180,957	493,975	3,399,811	962,095	5,026,564	1,376,095
宮川早生	2,204,581	499,010	4,493,711	1,125,907	6,772,031	1,620,052
興津早生	2,211,618	498,929	3,484,873	978,410	5,158,226	1,398,199
南柑20号	2,180,277	493,053	3,377,746	956,997	5,001,361	1,370,862
石地	2,197,774	495,991	3,423,911	967,803	5,071,470	1,383,658
青島	2,164,291	487,676	3,353,512	951,265	4,978,932	1,366,112
甘平	3,631,571	750,957	4,699,485	1,203,304	6,775,159	1,625,231
愛媛果試第28号	3,870,977	823,803	4,508,007	1,188,867	6,547,759	1,570,739
せとか	3,856,235	763,555	3,511,598	945,623	6,285,213	1,499,373
清見	3,851,969	811,723	3,861,103	1,041,701	6,224,234	1,498,544
スイートオレンジ	3,408,894	898,650	2,624,676	886,365	4,156,802	1,262,856
ワシントンネーブル	5,349,957	1,062,091	3,834,141	1,060,600	6,704,014	1,584,971
トロビタオレンジ	3,959,353	929,552	2,958,119	912,502	4,926,093	1,308,390
ダンカン	5,059,024	1,128,655	4,012,213	1,146,572	4,447,332	1,169,500
河内晩柑	5,825,461	1,167,535	4,992,221	1,293,377	4,601,370	1,142,759
麻豆ブンタン	4,168,370	1,006,183	3,449,374	1,014,763	1,828,694	482,888
Hunan pummelo	3,025,780	900,556	2,614,365	902,581	1,558,308	472,188
Tahitian pummelo	3,716,178	960,857	3,167,995	944,454	1,754,830	453,338
カオパン	3,143,981	907,822	2,624,106	896,037	1,421,368	437,875

# 遺伝子ごとに見てみると

参照配列 AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC

ポンカン AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC  
AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC

ウンシュウミカン AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC  
AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC

甘平 AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC

ワシントンネーブル AAAAAAAAAAGGGGGGGGGGTTTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC

河内晩柑 AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC

麻豆ブント AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGTT**A**TTTTTTTTCCCCCCCCC

# 遺伝子のタイプ別に分類

参照配列

AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC

ポンカン

AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC  
AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC

ウンシュウミカン

AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC  
AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC

甘平

AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC

ワシントンネーブル

AAAAAAAAAAGGGGGGGGGGGTTTTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC

河内晩柑

AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC

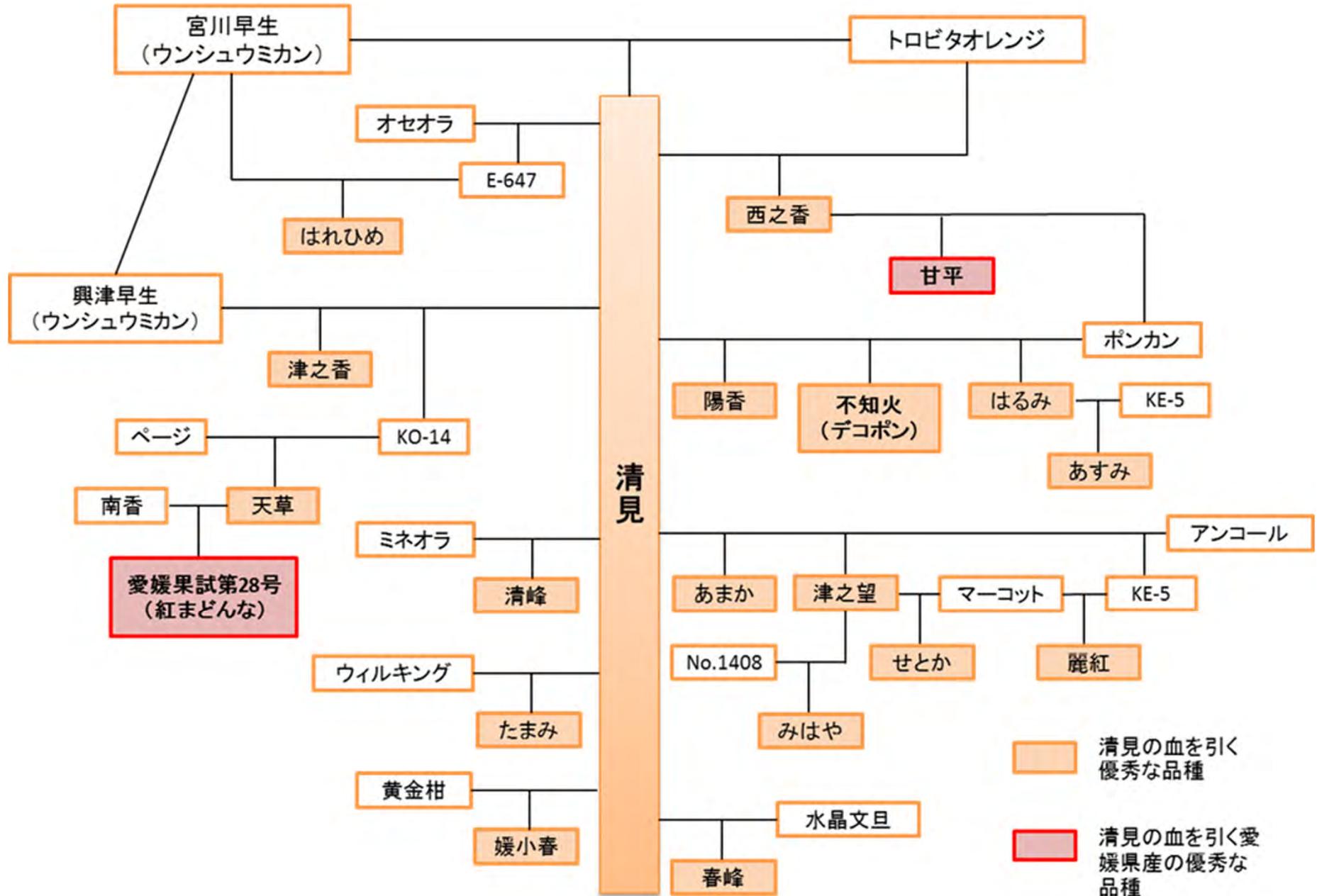
麻豆ブント

AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC  
AAAAAAA**T**AAGGGGGGGGGGGT**A**TTTTTTTTCCCCCCCCC

# 中晩柑の遺伝子タイプ

	甘平	愛果28	せとか	清見
ミカン型（複数のミカンで共通）	15397	12035	14452	10729
ミカン・オレンジ型（複数のミカンとオレンジで共通）	934	1088	1202	1557
ダンシー型	60	59	77	17
無核紀州型	42	79	61	53
クレオパトラ型	90	94	87	19
ポンカン型	209	96	91	43
クレメンティン型	216	484	209	138
マーコット型	41	94	145	52
ウンシュウミカン型	1671	1698	834	1244
ウンシュウミカン・オレンジ型	937	1371	632	1657
オレンジ型	1084	1862	1334	2318
ダンカン型（グレープフルーツ）	49	164	76	106
河内晩柑型	59	312	130	223
ブンタン型	380	974	1844	2033

# 最近の品種の系統図



# 各品種の特徴

	甘平	愛果28	せとか	清見
ミカン型（複数のミカンで共通）	<b>15397</b>	12035	14452	<b>10729</b>
ミカン・オレンジ型（複数のミカンとオレンジで共通）	934	1088	1202	1557
ダンシー型	60	59	77	17
無核紀州型	42	79	61	53
クレオパトラ型	90	94	87	19
ポンカン型	<b>209</b>	96	91	43
クレメンティン型	216	<b>484</b>	209	138
マーコット型	41	94	<b>145</b>	52
ウンシュウミカン型	1671	1698	834	1244
ウンシュウミカン・オレンジ型	937	1371	632	1657
オレンジ型	1084	1862	1334	<b>2318</b>
ダンカン型（グレープフルーツ）	49	164	76	106
河内晩柑型	59	312	130	223
ブンタン型	<b>380</b>	<b>974</b>	1844	<b>2033</b>

# これで何がわかるのか

- 品種間で共通な遺伝子 = 有用な形質
  - 品種特異的な遺伝子 = その品種の特徴を示す形質
    - ↓
    - 他の品種にはない優れた形質  
(食味や外観, 香りなど)
    - その品種で問題になっている形質  
(病虫害や生理障害など)
    - ↓
- 品種改良やゲノム編集に利用