

世界における植物ゲノム解析の現状と課題

○原田大士朗¹, 市原寿子², 中谷明弘², Ghelfi Andrea¹, 藤代継一¹, 小原光代¹, 平川英樹¹, 田畑哲之¹, 磯部祥子¹

1) かずさDNA研究所 2) 大阪大学 大学院医学系研究科

近年、シーケンシング技術の発展に伴い安価で効率的にゲノム解析がおこなえるようになった。モデル植物以外の種でも日々新たにゲノム解析がおこなわれており、2018年上半期だけでも約40種の植物において新規にゲノム解析がなされた。現時点で340種以上の植物ゲノムが解析されており、今後も新たな植物ゲノム情報が加速的に明らかになると考えられる。

このように植物ゲノム情報は世界的に拡充してきているが、いまだ課題も多い。本発表では植物ゲノム情報をキュレーションするうえで見てきた植物ゲノム解析の現状と課題を紹介する。

植物統合ポータルサイトPlant GARDEN

2019年春β版公開予定

Plant GARDEN

Plant GARDEN (Genome And Resource Database Entry) はさまざまな植物のゲノムやマーカー情報を集めたポータルサイトです

植物種からさがす

他の方法でさがす

解析してみよう

データ一覧

植物種からさがす

オランダダイダイ

キク

シロイヌナズナ

ダイコン

ダイズ

トマト

ミヤコグサ

ヨーロッパブドウ

ラッカセイ

ナンキンマメ

すべてみる ▶

形質から

配列から

キーワード検索

SNP Detection

その他の解析ツール

全情報

マーカー

形質

検索方法 ▶

- 簡潔な検索方法
- 操作を迷わせない
- 4つのコンテンツを整備

- 想定するユーザー
- 生物研究者
 - 育種関係者
 - 教育機関

植物ゲノム、トランスクリプトーム、プロテオーム、メタボローム、マーカー、変異、形質などの情報を提供。国内外の植物DBを統合し横断検索が可能。

PlantGARDENで公開予定の植物種

よく検索される植物種

オランダダイダイ

キク

シロイヌナズナ

ダイコン

ダイズ

トマト

ミヤコグサ

ヨーロッパブドウ

ラッカセイ

ナンキンマメ

イネ

オオムギ

キャッサバ

キャベツ

キュウリ

トウモロコシ

ナス

ハクサイ

メロ

全植物種リスト

学名 ↓

和名 ↓

Actinidia chinensis

キウイフルーツ

Aegilops tauschii

調査中

Aethionema arabicum

調査中

Allium fistulosum

ネギ

Amaranthus hypochondriacus

調査中

Plant GARDEN

検索方法 ▶

用語検索 ▶

JA EN

トマト / Solanum lycopersicum

配列名: SL3.0

ゲノム配列の概要

配列名	assembly_version
異称名	biological_sample_id
配列数	scaffold_count
シードモジュールの数	pseudomolecule_count
配列長 (bp)	assembly_size
N50長 (bp)	scaffold_N50
シーケンシングの方法	sequencing_method
取得した配列量	sequencing_depth
アセンブリ方法	assembly_method
推定ゲノムサイズ	estimated_genome_size
DOI	DOI
コメント1	sequencing_method_comments
コメント2	other_comments

現在ゲノム情報が公開されているほぼ全ての植物種を網羅した世界最大規模の統合情報サイトを目指す。

世界における植物ゲノム解析状況

plabi dataBase

https://www.plabipd.de/

Total 366

Curated 319

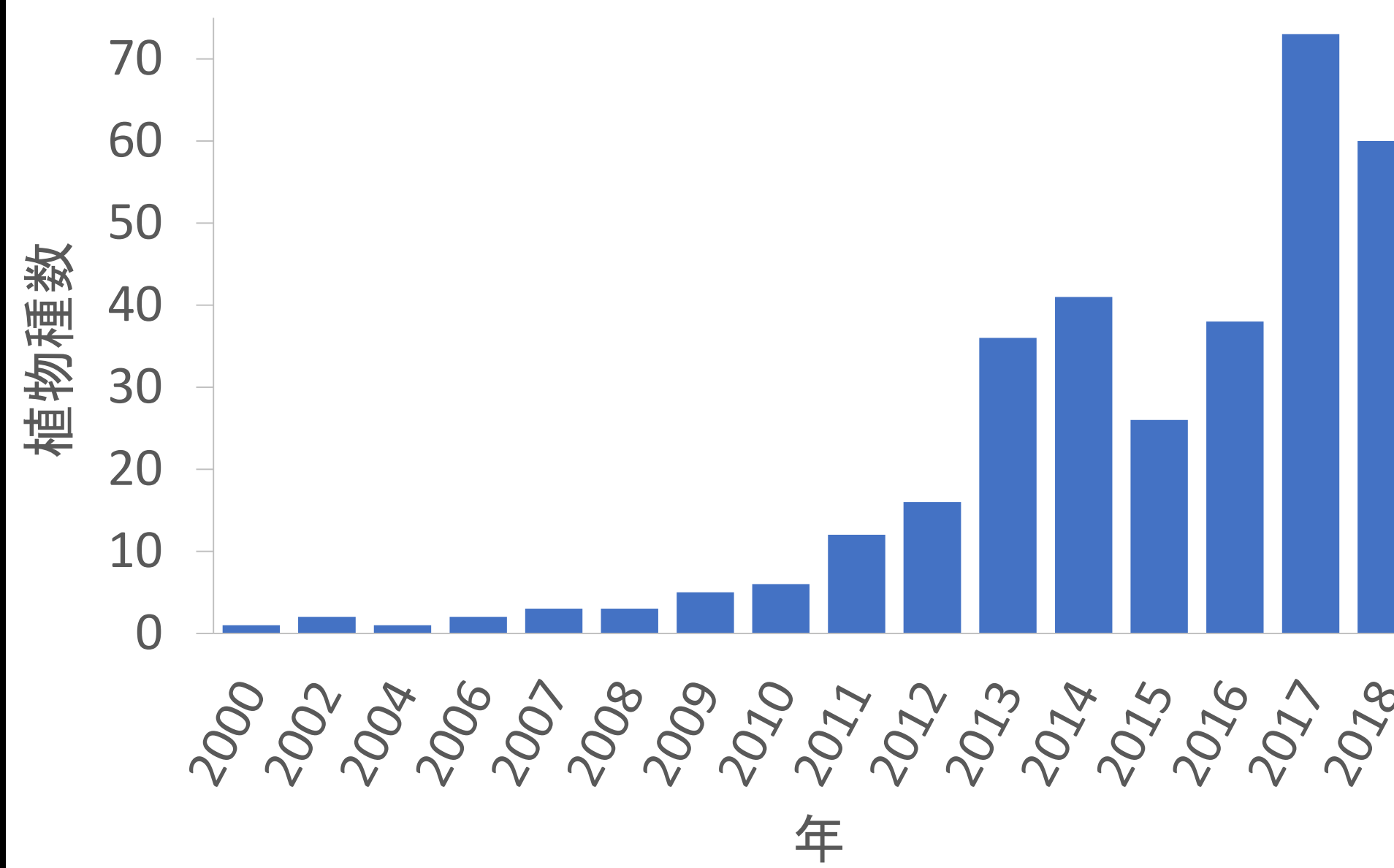
Plant genomes / Timeline view

2017

2018

植物種数

ゲノム解析がおこなわれた植物種数



シーケンシング技術の進歩

歩にとともに、ゲノム解析件数は年々増加傾向にある。2018年8月時点で360種以上の植物についてゲノム解析結果が報告された。しかし、約50種の植物種についてはゲノム情報のキュレーションが不可能だった。

ゲノム情報の品質

ゲノムアセンブリーレベルと付属ファイル数

Assembly Level	CDS	CDS, GFF, AA_fasta
Complete Genome	3	3
Chromosome	89	52
Scaffold	138	43
Contig	19	6
unknown	20	12

各アセンブリーレベルのばらつき

cluster	%			
	Ch	Sca	Con	Unk
1-1	10	1	0	0
1-2	12	15	18	0
1-3	10	7	5	0
1-4	10	56	50	0
1-5	0	10	5	43
1-6	9	0	14	0
1-7	0	2	5	0
2	41	4	5	14
3	3	0	0	14
4	5	0	0	0
5	0	1	0	0
6	0	3	0	29
7	0	1	0	0

キュレーション可能だった植物ゲノム情報のうち、Pseudomolecule以上の植物は92種だった。

Pseudomolecule以上の植物中、CDS, GFF, AA fastaが揃っている種は55種であった。

Chromosomeレベルでアセンブリされた植物でも、アセンブリの結果にはばらつきが大きかった。また、Scaffoldレベルであっても長くアセンブリできている植物種も存在していた。

Ward's method
X: N50 Value,
Y: Scaffold or Contig Counts

本プロジェクトは、JSTバイオサイエンスデータベースセンター(NBDC)「統合化推進プログラム」の支援を受けた。

