Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Dario Cantu
M. Andrew Walker *Editors*

The Grape Genome

Springer

Compendium of Plant Genomes

Series Editor

Chittaranjan Kole, Raja Ramanna Fellow, Government of India, ICAR-National Research Center on Plant Biotechnology, Pusa, New Delhi, India

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 100 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant?

Please contact Prof. C. Kole, Series Editor, at ckoleorg@gmail.com

More information about this series at http://www.springer.com/series/11805

Dario Cantu · M. Andrew Walker
Editors

The Grape Genome

Springer

Editors
Dario Cantu
Department of Viticulture and Enology
University of California, Davis
Davis, CA, USA

M. Andrew Walker
Department of Viticulture and Enology
University of California, Davis
Davis, CA, USA

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-18600-5 ISBN 978-3-030-18601-2 (eBook)
https://doi.org/10.1007/978-3-030-18601-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book series is dedicated to my wife Phullara, and our children Sourav, and Devleena
Chittaranjan Kole

Harold Olmo (left) and Al Koyama (center), his grape breeding assistant of many years, and Andy Walker (right) under the Winkler Vine in the UC Davis vineyards in 2003 (Picture by Daniel Ng)

This book is dedicated to the memory of Harold P. Olmo. He was the leading figure in grape genetics and breeding for 40 years and had a remarkable influence on viticulture across the globe. His extensive travels (by car, train, foot, and horse) through Afghanistan and Iran collecting grapes, Prunus *and other horticultural crops while avoiding disasters, gunshots, angry tribal disputes, earned him the moniker "The Indiana Jones of Viticulture". He released wine grapes, table grapes, raisin grapes and rootstocks, and was an excellent ampelographer. May his inspirational viticultural spirit live on.*

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of "markers" physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F2 were utilized, and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in the studies of evolution and phylogenetic relationship, genetic diversity, DNA fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still, they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated the development of the "genomic resources" including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic–physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, the emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, the sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But the development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics, and a new subject was born—bioinformatics.

Thus, the evolution of the concepts, strategies, and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker, and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second-generation sequencing methods. The development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant Arabidopsis thaliana in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, the development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series "Compendium of Plant Genomes," a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and three basal plants is accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization are growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is of interest not only for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology,

physiology, pathology, entomology, nematology, crop production, biochemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are, therefore, focusing on the basic aspects of the genomes and their utility. They include information on the academic and/or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, the potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor, it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with a lifetime experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to Springer staff particularly, Dr. Christina Eckey and Dr. Jutta Lindenborn, for the earlier set of volumes and presently Ing. Zuzana Bernhart for all their timely help and support.

I always had to set aside additional hours to edit books beside my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

New Delhi, India Chittaranjan Kole

Preface

Grapevines (*Vitis vinifera*) have been a source of food and wine since their domestication nearly 8000 years ago. Grape is one of the most important horticultural crops in the world, with over 7 million hectares planted worldwide. In addition to its economic value, grapevine is a model organism for the study of perennial fruit crops and non-climacteric fruit ripening. Its economic and scientific importance made *V. vinifera* an obvious early candidate for genomic sequencing. The two draft genome references released in 2007 were the second publicly available genomes of a woody species and the fourth of a flowering plant. The genome assembly of the experimental inbred line released by "The French–Italian Public Consortium for Grapevine Genome Characterization," PN40024, has served as reference for thousands of genetic and transcriptomic studies. Now over a decade since its release, the PN40024 genome is still a valuable resource to the grapevine community thanks to the continuous effort of the Consortium to improve its structure and annotation.

However, it was understood that a single reference genome was inadequate for studying the function of non-reference cultivar genomes. Seminal work in Tannat and other wine grape cultivars showed substantial unshared gene content between grape cultivars. Recent advancements in sequencing technologies and bioinformatics have made it feasible to generate genome references for other cultivars of equivalent or greater quality than that of PN40024. The genome assemblies of Cabernet Sauvignon, Chardonnay, Carménère, and Zinfandel were released in the last two years. A *V. riparia* genome assembly was released when this book was in the final stages of production; we expect many more genome references for *Vitis* species to be publicly available in the next few years, including those of North American and Asian accessions that are being produced in our laboratories as part of National Science Foundation (1741627) and USDA National Institute of Food and Agriculture (2017-51181-26829) projects. Our research groups have been contributing to the recent advancements in *V. vinifera* genomics. This has been possible because of support from E. & J. Gallo Winery, J. Lohr Vineyards and Wines, Dolce Winery, the Louis P. Martini Endowment in Viticulture, Viña San Pedro, Concha y Toro, UC Davis Chile Life Sciences Innovation Center, and the Chilean Economic Development Agency, and the collaboration between our groups and the scientists at Pacific Biosciences, specifically Paul Peluso, Jason Chin, David Rank, Kristin Mars, and Emily Hatas.

Today, grape cultivation, sustainability, and security rely heavily on North American *Vitis* species as sources of resistance to abiotic and biotic stresses. This reliance originated in the 1860s when the European wine industry was saved by the use of North American species as rootstocks. Currently, more than a dozen North American and Central Asian varieties are used in breeding programs as sources of resistance to abiotic and biotic stresses, either for rootstocks or hybridized with *V. vinifera* for the scion. We expect that genetic diversity, breeding, and biotechnology will play a critical role for sustaining viticulture when faced with a changing climate and other challenges as they arise.

The sixteen chapters of this volume provide a comprehensive review of early and ongoing efforts to discern the genetics, genomics, and breeding of the grapevine. We are grateful to all the authors for their contributions. We would like to thank Prof. Chittaranajan Kole, Editor-in-Chief of the Genome Compendium Series, for inviting us to contribute this volume as well as Naresh Kumar Mani, Manopriya Saravana, and the staff at Springer for their help. We would also like to thank Jadran Francisco Garcia Navarrete, Mélanie Massonnet, Rosa Figueroa-Balderas, Amanda Vondras, and Summaira Riaz for helping review and edit the chapters. Dario would also like to thank his wife, Annegret, and daughters, Amanda and Adele, for their infinite patience and support during the two-year journey that turned an idea into a table of contents and finally into a book.

Davis, USA

Dario Cantu
M. Andrew Walker

Contents

Abbreviations

(s)PLS	(Sparse) Partial least square regression
1-MCP	1-Methylcyclopropene
2,4-D	2,4-Dichlorophenoxyacetic acid
2-CEPA	2-Chloroethylphosphonic acid
2-DE	Two-dimensional electrophoresis
4CL	4-Coumarate-CoA ligase
5mC	5 Methylcytosine
AB	Advanced backcross
ABA	Abscisic acid
ABC	ATP-binding cassette
ABF	Abscisic acid response element-binding factor
AB-QTL	Advanced backcross QTL
AC	After Christ
ACC	1-Aminocyclopropane-1-carboxylic acid
ACO	1-Aminocyclopropane-1-carboxylic acid oxidase
ACR	AC-rich
ACS	ACC synthase
AD	Anno Domini
aDNA	Ancient DNA
ADP	Adenosine diphosphate
AFLP	Amplified fragment length polymorphism
AGAP	Amélioration génétique et adaptation des plantes méditerranéennes et tropicales
AI	Acidic invertases
AIL	AINTEGUMENTA-like
AM	Association mapping
amiRNAs	Artificial miRNAs
ANR	Anthocyanidin reductase
ANT	AINTEGUMENTA
AOC	Appellation d'Origine Contrôlée
AOS	Allene oxide synthase
AP2/ERF	APETALA 2/ethylene-responsive element-binding factor
APHIS	Animal and Plant Health Inspection Service
APT	Adenine phosphoribosyl transferase
AQUILO	AcQUIred tolerance to LOw temperatures

AraNet	Probabilistic functional gene network of Arabidopsis thaliana
ARF	Auxin response factor
ARS	Agricultural Research Service
ATAC-seq	Assay for transposase accessible chromatin sequencing
ATP	Adenosine triphosphate
ATRX	Arabidopsis trithorax-related protein
AUDPC	Area under the disease progress curve
AuxRE	Auxin response elements
AVA	American viticultural areas
AVG	Aminoethoxyvinylglycine
B	Billions
BAC	Bacterial artificial chromosome
BAH	Bromo-adjacent homology
BAP	6-Benzylaminopurine
BC	Before Christ
BCE	Before Common Era
BeYDV	*Bean yellow dwarf virus*
Bgh	*Blumeria graminis* f. sp. *hordeii*
BGs	β-Glucosidases
BLAST	Basic local alignment search tool
BOINC	Berkeley Open Infrastructure for Network Computing
BR	Brassinosteroid
BSA	Bulked segregant analysis
BUSCO	Benchmarking universal single-copy orthologs
bZIP	Basic leucine zipper domain
C4H	Cinnamate-4-hydroxylase
CAT	Chloramphenicol acetyltransferase
CC	Coiled coil
CCA	Canonical correlation analysis
CCoAOMT	Caffeoyl-CoA 3-O-methyltransferase
cDNA	Complementary DNA
cDNA-AFLP	cDNA-amplified fragment length polymorphism
CDS	Coding sequence
CEC	Cation exchange capacity
CG(s)	Candidate gene(s)
ChIP-seq	Chromatin immunoprecipitation sequencing
CHK	Cytokinin histidine kinase
Chr	Chromosome
CHS	Chalcone synthase
CIMIS	California Irrigation Management Information System
CIRAD	Centre de coopération internationale en recherche agronomique pour le développement
CKX	Cytokinin oxidase/dehydrogenase

CLF	Curly leaf
cM	CentiMorgans
CMT(s)	Chromomethylase(s)
CNR	Colorless non-ripening
CO_2	Carbon dioxide
COI	Coronatine insensitive
COLOMBOS	COLlection Of Microarrays for Bacterial OrganismS
COMT	Caffeic acid 3-O-methyltransferase
CORFO	Chilean Economic Development Agency
COST	European Cooperation in Science and Technology
CP	Coat protein
CRE(s)	Cis-regulatory element(s)
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/Cas9-associated protein
CT	Computed tomography
CTAB	Cetyltrimethylammonium bromide
DAP-seq	DNA-affinity-purified sequencing
DART-MS	Direct analysis in real-time-mass spectrometry
DB(s)	Database(s)
DCL	Dicer-like ribonuclease III
DEF	Deficiens
DFR	Dihydroflavonol reductase
DGE	Differentially expressed gene
DM	Downy mildew
DME	Demeter
DML	Demeter-like protein
DMR(s)	Differentially methylated region(s)
DNA GL(s)	DNA glycosylase lyase(s)
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DPA	Diphenylamine
dpi	Days post-inoculation
DR	Decamers
DRM(s)	Domain rearranged methyltransferase(s)
dsRNA(s)	Double-strand RNA(s)
EBI	European Bioinformatics Institute
ELISA	Enzyme-linked immunosorbent assays
ELIXIR-EXCELERATE	European life sciences infrastructure for biological information
EMBL	European Molecular Biology Laboratory
EMPHASIS	The European Infrastructure for Multi-Scale Plant Phenomics and Simulation
ENCODE	Encyclopedia of DNA elements
eQTL	Expression QTL
ERF(s)	Ethylene response factor(s)
ESC	Extra sex comb

ESFRI	European Strategy Forum on Research Infrastructures
ESI	Electrospray ionization
EST(s)	Expressed sequence tag(s)
ETI	Effector-triggered immunity
eTM	Endogenous target mimics
ETR	EThylene receptor
EZ	Enhancer of zeste
F	Female
F3'5'Hs	Flavonoid-3',5'-hydroxylases
F3H	Flavanone 3-hydroxylase
F3'Hs	Flavonoid-3'-hydroxylases
FAIR	Findability, accessibility, interoperability, and reusability
FAO	Food and Agricultural Organization of the United Nations
FAOSTAT	Food and Agriculture Organization Corporate Statistical Database
FAS	Fatty acid synthase
FD	Flavescence dorée
FDp	FD-phytoplasma
FEELnc	Flexible extraction of long non-coding RNA software
FEM	Fondazione Edmund Mach
FIE	Fertilization-independent endosperm
flb	Fleshless berry mutation
FLC	Flowering Locus C
Flp	Flippase
FLS	Flavonol synthases
FLS(s)	Flavonol synthase(s)
FRT	Flippase recognition target
fruitENCODE	Fruit encyclopedia of DNA elements
FT	Flowering Locus T
FUL	Fruitfull
FUM	Fumarase
GA(s)	Gibberellin(s)
GABA	Gamma-aminobutyric acid
GAox	GA-oxidases
GbM	Gene body methylation
Gbp	Gygabase pairs
GBS	Genotyping by sequencing
GC	Gas chromatography
GC-MS	Gas chromatography–mass spectrometry
GCN(s)	Gene co-expression network(s)
gDNA	Genomic DNA
GENCODE	Genomic encyclopedia of DNA elements
GEO	Gene expression omnibus

GFF	General feature format
GFLV	Grapevine fanleaf virus
GFP	Green fluorescent protein
GL	Glycosylase lyases
GLRaV	Grapevine leafroll-associated virus
g_m	Mesophyll conductance
GMO	Genetically modified organism
GPS	Global Positioning System
GrapeIS	Grape Information System
GRBaV	Grapevine red blotch-associated virus
gRNA	Guide RNA
GRSPaV	Grapevine rupestris stem pitting-associated virus
GS	Genomic selection
g_s	Stomatal conductance
GS/MS	Gas chromatography–mass spectrometry
GST	Glutathione-S-transferase
GTD(s)	Grapevine trunk disease(s)
GUS	Beta-glucuronidase
GWAS	Genome-wide association scans/studies
GxE	Genotype by environment interaction
H	Hermaphrodite
H_2O_2	Hydrogen peroxide
HAT(s)	Histone acetyltransferase(s)
HB	HD-Zip homeobox
HDAC(s)	Histone deacetylase(s)
HDMT(s)	Histone demethylase(s)
HDP1	Harbinger transposon-derived protein 1
HGAP	Hierarchical genome assembly process
Hi-C	Genome-wide chromatin conformation capture protocol
HMT(s)	Histone methyl transferase(s)
HMW	High molecular weight
HPLC	High-performance liquid chromatography
HPLC-MS	High-performance liquid chromatography-mass spectrometry
HPTM(s)	Histone post-translational modification(s)
HR	Hypersensitive response
HRM	High-resolution melting analysis
HS	Headspace
HT	High-throughput
HTML	Hypertext Markup Language
HTs	Hexose transporters
HTS	High-throughput sequencing
HY5	Elongated hypocotyl 5
HYH	HY5 homologue
IAA	Indole-3-acetic acid
IBMP	3-Isobutyl-2-methoxypyrazineare

ICP-MS	Inductively coupled plasma mass spectrometry
IDM1	Increase in DNA methylation 1
IGGP	International Grape Genome Program
indels	Single-base insertions or deletions
INTEGRAPE	Data integration to maximize the power of omics for grapevine improvement
IPCC	Intergovernmental Panel on Climate Change
IPMP	3-Isopropyl-2-methoxypyrazine
IPT	Isopentenyltransferase
IR	Infrared light
Iso-Seq	Isoform sequencing
IT	Information technologies
iTRAQ	Isobaric tags for relative and absolute quantitation
ITS	Internal transcribed spacer region
JA	Jasmonic acid
JA-Ile	Jasmonic acid—isoleucine
JAZ	Jasmonate ZIM domain
JGI	Joint Genomics Institute
JMJ12	Jumonji domain-containing protein 12
JMT	S-adenosyl-l-methionine:jasmonic acid carboxyl methyltransferase
K Ha	Thousand hectares
K	Potassium
kb	Kilobases
K_{leaf}	Leaf hydraulic conductance
KT	Thousand tons
KYP	Kryptonite
LAR	Leucoanthocyanidin reductases
LB	Left T-DNA border
LBD	Lateral Organ Boundaries Domain
LC	Liquid chromatography
LC-DAD	LC-Diode array detector
LD	Linkage disequilibrium
LDOX	Leucoanthocyanidin dioxygenase
LGN(s)	Local gene network(s)
LiCl	Lithium chloride
L-IdnDH	L-Idonate dehydrogenase
lncRNA(s)	Long non-coding RNA(s)
LOB	Lateral Organ Boundaries Domain family
LOD	Logarithm of the odds
LOG	Phosphoribohydrolase ‘Lonely Guy’
LOX	Lipoxygenase
Lp_r	Root hydraulic conductivity
LRR	Leucine-rich repeat domain
LTQ	Linear trap quadrupole
M	Male
MAPK(s)	Mitogen-activated protein kinase(s)

MAS	Marker-assisted selection
Mb	Megabases
MBD7	Methyl CpG-binding protein 7
MDH	Malate dehydrogenase
MDS	Multidimensional scaling
ME	Malic enzyme
MEA	Medea
MeDIP-seq	Methyl DNA immunoprecipitation sequencing
MeJA	Methyl jasmonate
MEMS	Methylation monitoring sequence
MEP	2C-Methyl-D-erythritol-4-phosphate
MET1	Methyltransferase 1
MFA	Multiple factor analysis
Mg	Magnesium
MH	Million hectares
microCT	Micro-computed tomography
MIP	Major intrinsic protein family
miRNA	MicroRNA
ML	Maximum likelihood
MLO	Mildew resistance Locus O
MML	Modified maximum likelihood
MNase-seq	Micrococcal nuclease sequencing
MP	Movement protein
Mpa	Megapascal
MPs	Methoxypyrazines
MRG	Morf-related gene
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MSAP(s)	Methylation-sensitive amplification polymorphism(s)
MSI1	Multicopy suppressor of IRA 1 protein
MT	Million tons
mtDNA	Mitochondrial DNA
MVA	Cytosolic mevalonate
My	Million years
Mya	Million years ago
NaCl	Sodium chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
NaOAc	Sodium acetate
NB	Nucleotide-binding site
NBT(s)	New breeding technique(s)
NCBI	National Center for Biotechnology Information
NCED	9-cis-epoxycarotenoid dioxygenase
ncRNA	Non-coding RNA
N_e	Effective population size

NES^2RA	Network expansion by stratified variable subsetting and ranking aggregation
ng	Nanograms
NGO	Non-governmental organization
NGS	Next-generation sequencing
NI	Neutral invertases
NMR	Nuclear magnetic resonance
NO	Nitric oxide
NOA	Naphthoxyacetic acid
NOR	Non-ripening
NSF PGRP	National Science Foundation: Plant Genome Research Program
nuDNA	Nuclear DNA
O_2^-	Oxygen
OIV	International Organization of Vine and Wine
OPR	12-Oxophytodienoate reductase
ORCAE	Online Resource for Community Annotation of Eukaryotes
P5CS	1-Pyrroline-5-carboxylate synthetase
PA	Polyamide
PAL	Phenylalanine ammonia lyase
PAM	Protospacer adjacent motif
PBA	Pedigree-based analysis
PCA	Principal component analysis
PCD	Programmed cell death
PcG	Polycom group proteins
PCR	Polymerase chain reaction
PD	Pierce's disease
PDH	Proline dehydrogenase
PDO	Protected designations of origin
PdR	PD resistance locus
PDS	Phytoene desaturase
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PFGE	Pulse field gel electrophoresis
PGDBj	Plant Genome DataBase Japan
PGIP	Polygalacturonase-inhibiting protein
PIP	Plasma membrane Intrinsic Protein
PlantGDB	Plant Genome Database
PLEXdb	Plant Expression Database
PM	Powdery mildew
PODC	The Plant Omics Data Center
PP2C	2C protein phosphatases
PR	Pathogenesis-related proteins
PRC2	Polycomb repressive complex 2
pre-miRNA	Precursor miRNA
pri-miRNA	Primary microRNA

PSII	Photosystem II
PTI	Pattern-triggered immunity
PTM(s)	Post-translational modification(s)
PVP	Polyvinylpyrrolidone
qPCR	Quantitative PCR
QqQ	Triple quadrupole
QTL(s)	Quantitative trait locus/loci
RAD	Restriction-site associated DNA
RAPD	Random amplification of polymorphic DNA
RB	Right T-DNA border
Rcg	Resistance to crown gall
Rda	Resistance to diaporthe ampelina
RdDM	RNA-directed DNA methylation
rDNA	Ribosomal DNA
Refseq	Reference sequence database
Ren	Resistance to erysiphe necator
RFLP	Restriction fragment length polymorphism
RFO(s)	Raffinose family of oligosaccharide(s)
RGAs	Resistance genes analogous
R-genes	Resistance genes
rin	Ripening inhibitor
RIN	RNA integrity number
RISC	RNA interference silencing complex
RK	Receptor kinase
RNA	Ribonucleic acid
RNAi	RNA interference
RNase A	Ribonuclease A
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
ROS1	Repressor of Silencing 1
Rpv	Resistance to plasmopara viticola
RR	Response regulators
RT-qPCR	Reverse transcription quantitative PCR
RUBISCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
Run	Resistance to uncinula necator
SAR	Systemic acquired resistance
SAS	Statistical analysis Software
SBP	SQUAMOSA promoter-binding protein
SBP-box/SPL	SQUAMOSA promoter-binding protein-like transcription factor
SCAR	Sequence-characterized amplified region
SE	Somatic embryogenesis
siRNA	Small interfering RNA
SLAM	Simultaneous localization and mapping
SMC	Sequential Markovian coalescent

SMRT	Single-molecule real-time sequencing
sNCGGa	Super-Nomenclature Committee for Grape Gene Annotation
SNP(s)	Single-nucleotide polymorphism(s)
SnRK(s)	Serine/Threonine-protein kinase(s)
SPE	Solid-phase extractions
SPME	Solid-phase microextraction
SRA	SET- and RING-associated
sRNA(s)	Small RNA(s)
sRNA-Seq	Small RNA sequencing
S-SAP	Sequence-specific amplification polymorphism
SSCP	Single-strand conformation polymorphism
SSE	Sum of square errors
SSR(s)	Simple sequence repeat(s)
STSs	Stilbene synthase genes
Su(z)12	Suppressor of zeste 12
SUTs	Sucrose transporters
SUVH	Suppressor of variegation homologue
SWEET	Sugars will eventually be exported transporter
SWN	Swinger
T/Ha	Tons per hectare
TAA1/TAR	TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS1/TRYPTOPHAN AMINOTRANSFERASE RELATED
TAGL	Tomato agamous-like
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
T-DNA	Transfer DNA
TDZ	Thidiazuron
TE(s)	Transposable element(s)
TF(s)	Transcription factor(s)
TFBS(s)	Transcription factor-binding site(s)
TG	Translucent green
Ti	Tumor inducing
TILLING	Targeting-induced local lesions in genomes
TIP	Tonoplasm intrinsic protein
TIR	Toll/Interleukin-1 receptor
TM	Thompson Seedless
TMT	Tandem mass tagging
TPSs	Terpene synthases
TQD	Triple quadrupole
transPLANT	Trans-national infrastructure for plant genomic science
TS	Thompson Seedless
TSS	Transcriptional start site
TT8	Theban Tomb 8

TTB	Alcohol and Tobacco Tax and Trade Bureau (US Department of the Treasury)
UAS	Unmanned aerial systems
UFGT	UDP-glucose flavonoid-3-O-glucosyltransferase
UGTs	UDP-glucosyltransferases
UHPLC	Ultra-high-performance liquid chromatography
UNIPROT	Universal Protein Resource
UPD	Uridine diphosphate
URGI	Unité de Recherche Génomique Info
USDA	United States Department of Agriculture
USDA/FAS	United States Department of Agriculture: Foreign Agricultural Service
USDA/NASS	United States Department of Agriculture National Agricultural Statistics Service Information
UTR	Untranslated region
UV	Ultraviolet light
VESPUCCI	Vitis Expression Studies Platform Using COLOMBOS Compendia Instances
VIB	Vlaams Instituut voor Biotechnologie
VIGS	Virus-induced gene silencing
VIM	Variant in methylation
VOC(s)	Volatile organic compound(s)
VPD	Vapor pressure deficit
VR	*Vinifera* x *Rotundifolia*
VTCdb	ViTis Co-expression DataBase
WET	Wine Equalization Tax
WGBS	Whole-genome bisulfite sequencing
WT	Wild type
WUE	Water-use efficiency
ya	Years ago
Y_{leaf}	Leaf water potential
YUC	YUCCA genes
ZEP	Zeaxanthin epoxidase
ZF	Zinc finger
ZFN	Zinc finger nuclease

Grapes in the World Economy

1

Julian M. Alston and Olena Sambucci

Abstract

With a farm gate value in 2016 of US$68 billion, grapes are the world's third most valuable horticultural crop (after potatoes and tomatoes). Cultivation of grapes for fruit and wine began at least 7000 years ago in the Near East, and over the millennia, thousands of cultivars have been developed and selected for particular purposes. Nowadays, grapes are grown all around the world, but mainly in places having a temperate, Mediterranean-style climate, and they are used to produce diverse consumer products including wine, table grapes, raisins, grape juice concentrate and distillate for various industrial uses as well as making fortified wine and brandy. The cultivars of grapes used to make these diverse products are likewise diverse, but a relatively small number account for the vast majority of production in each major category. Predominantly, European *V. vinifera* scions are grown on rootstock from phylloxera-resistant Native American species. Particular cultivars are valuable to farmers in particular applications for their agronomic traits and fruit-quality traits, which together determine the value of the crop and the cost of producing it. These values can be conditioned by consumer preferences for attributes of the production process and by government policies including trade taxes, alcohol excise taxes, and regulations over production practices or limiting yields. Evolving demands for traits create demands for work by viticulturists and other scientists to understand the grape genome and work with it.

1.1 Grapes in the World Economy

Archeological evidence suggests stone-age people were making wine from grapes in Georgia and Armenia 8000 years ago, and grapes have been cultivated for winemaking for at least 7000 years (McGovern 2003)—well before the time of the "Epic of Gilgamesh," set in Mesopotamia around 2100 BCE, which is the first written account of grapes and wine. Over the millennia, and especially during the past 500 years, *Vitis vinifera* grapevines originating from the Near East have spread to all four corners of the world. Thousands of cultivars have

J. M. Alston (✉) · O. Sambucci
Agricultural and Resource Economics, University of California Davis, One Shields Ave, Davis, CA 95616, USA
e-mail: julian@primal.ucdavis.edu

O. Sambucci
e-mail: sambucci@primal.ucdavis.edu

D. Cantu and M. A. Walker (eds.), *The Grape Genome*, Compendium of Plant Genomes, https://doi.org/10.1007/978-3-030-18601-2_1

been generated and selected for particular purposes; and thousands more are known, including many wild varieties.[1]

Grapes are grown for diverse end uses, beyond wine production. *V. vinifera* grapes, along with non-*vinifera* varieties or hybrids, are eaten as fresh table grapes, dried to make raisins, or crushed either to produce grape juice concentrate, or to be fermented and distilled for industrial use as well as for use in making alcoholic beverages; and they are used as ornamental plants. These diverse end uses call for different varietal traits, and thus many diverse varieties, but a relatively small number account for the vast majority of production in each major category. Predominantly, European *V. vinifera* scions are grown on rootstock from phylloxera-resistant American species such as *Vitis aestivalis*, *rupestris*, and *riparia*. Although the genus includes a total of 79 "accepted" species (The Plant List: Vitis 2018), predominantly from North America and the Near East, the vast majority of today's cultivated grapes are varieties of *V. vinifera*, and only a few varieties from other species and some hybrids are of commercial significance.

Grapes are significant in the global economy. In 2016, the world produced 77.4 million tonnes (MT) of grapes (worth some $68.3 billion at the farm) from 7.1 million hectares (MH) of vineyard—a 50 percent increase over the 52.0 MT produced from 9.5 MH in 1966. These grapes are used to produce food and wine at retail worth several times the farm value of the grapes themselves. Over the 50 years, 1966–2016, global average yields almost doubled, from 5.5 to 10.9 tonnes per hectare (T/Ha), and the farm value of grape production grew from $29.6 billion to $44.3 billion in real (2004–2006 international dollar) terms, even though the total vineyard area shrank by one-quarter.[2] Changes in grape cultivars contributed directly to the growth in yield, production, and economic value, and while many other aspects of grape production also changed—including where in the world grapes are grown, how, and for which end uses—these aspects are all chosen jointly with varieties.

Looking to the future, the demand for genetic innovation in grape production will depend in part on the patterns of growth in demand for grape products. Growth in population and per capita incomes would be expected to cause an increase in demand for all grape products, with a relative increase in the demand for more income-elastic fresh versus dried grapes and premium versus more basic wine. Where that growth is to take place around the world will matter, too. In the context of a market driven by broad shifts in final consumer demand, growers will continue to demand cultivars of scions (and rootstocks) that produce fruit with desired quality attributes and have desired agronomic attributes: higher yielding, resistant to pests and diseases, and tolerant of environmental stresses.

This chapter provides an introductory overview of the economic geography (and, where relevant, economic history) of the cultivation of grapes around the world with an eye to how these aspects relate to the grape genome, which is the broader subject of the volume. We discuss the patterns of production of grapes for each of the main end uses, and how they have been changing, and the roles of genetic traits of cultivars as contributors to those patterns. We consider the value of particular traits to producers in specific settings and how these values are influenced by evolving market demand for product and process attributes of food and beverage products, government policy as a conditioning factor, and the changing natural environment, including the ever-present and evolving pests and diseases and, more recently, climate. The chapter begins with an overview of grape production around the world in terms of where grapes are grown, and recent trends in production and utilization.

[1] In the preface to their book describing 1368 varieties of wine grapes, Robinson, Harding, and Vouillamoz (2012, p. viii) suggest the "total number of different vine varieties is about 10,000."

[2] Statistics reported in this section are based primarily on FAOSTAT (2018); Table 1.1 includes more detailed data for 2016.

Table 1.1 Area, volume, yield, and value of grape production in 2016, by regions and countries

Region and country	Total area (K Ha)	Volume (KT)	Yield (T/Ha)	Value ($ M)	Average unit value ($/T)
Africa	**349.6**	**4882.5**	**14.0**	**3463.7**	**709**
Egypt	74.9	1716.8	22.9	567.9	331
South Africa	120.5	2008.8	16.7	1780.1	886
Americas	**1001.4**	**13,659.4**	**13.6**	**12,747.5**	**933**
Argentina	223.9	1758.4	7.9	358.7	204
Brazil	77.0	984.5	12.8	596.6	606
Chile	203.1	2473.6	12.2	4455.0	1801
Peru	27.9	690.0	24.7	490.9	711
North America	**421.9**	**7188.6**	**17.0**	**5236.8**	**728**
USA	409.9	7097.7	17.3	5130.3	723
Asia	**2122.6**	**28,918.4**	**13.6**	**22,249.9**	**769**
Uzbekistan	135.1	1642.3	12.2	489.4	298
China and HK	843.4	14,842.7	17.6	14,007.2	944
Afghanistan	82.5	874.5	10.6	*392.7*	*449*
India	122.0	2590.0	21.2	1837.1	709
Iran	207.3	2450.0	11.8	801.8	327
Turkey	435.2	4000.0	9.2	1967.3	492
Europe	**3446.9**	**27,797.1**	**8.1**	**28,325.3**	**1019**
Romania	175.1	736.9	4.2	523.9	711
Greece	112.3	990.3	8.8	771.3	779
Italy	668.1	8201.9	12.3	3311.9	404
Portugal	175.0	773.9	4.4	1463.6	1891
Spain	920.1	5934.2	6.4	4487.9	756
France	757.2	6247.0	8.2	14,496.1	2320
Germany	100.0	1225.6	12.3	1298.3	1059
Oceania	**176.4**	**2181.4**	**12.4**	**1506.4**	**691**
Australia	136.3	1772.9	13.0	991.1	559
World total	**7096.7**	**77,438.9**	**10.9**	**68,292.9**	**882**

Notes Value and average unit value for Afghanistan (in italics) calculated as weighted averages for the region
Sources Created by the authors using data from FAOSTAT (2018) and USDA/FAS (2018a)

1.1.1 Grape Production and Utilization

Table 1.1 and Fig. 1.1 provide statistics on the production of grapes around the world in terms of area of vineyard, average yield, production, total value of production, and average unit value, drawing on data from FAOSTAT (2018).[3] In 2016, the world had a total of 7.1 MH planted to grapes. Five countries (Spain, China, France, Italy, and Turkey) accounted for 3.6 MH, just over half the total area, and just 15 countries accounted for 5.5 MH, more than three-quarters.

[3]We draw on various sources for data, including the International Organization of Vine and Wine (OIV), the Food and Agricultural Organization of the United Nations (FAO), the United States Department of Agriculture Foreign Agriculture Service (USDA/FAS), Anderson and Aryal (2013), and Anderson and Pinilla (2018). The Appendix provides more detailed data tables and some discussion of the different data sources.

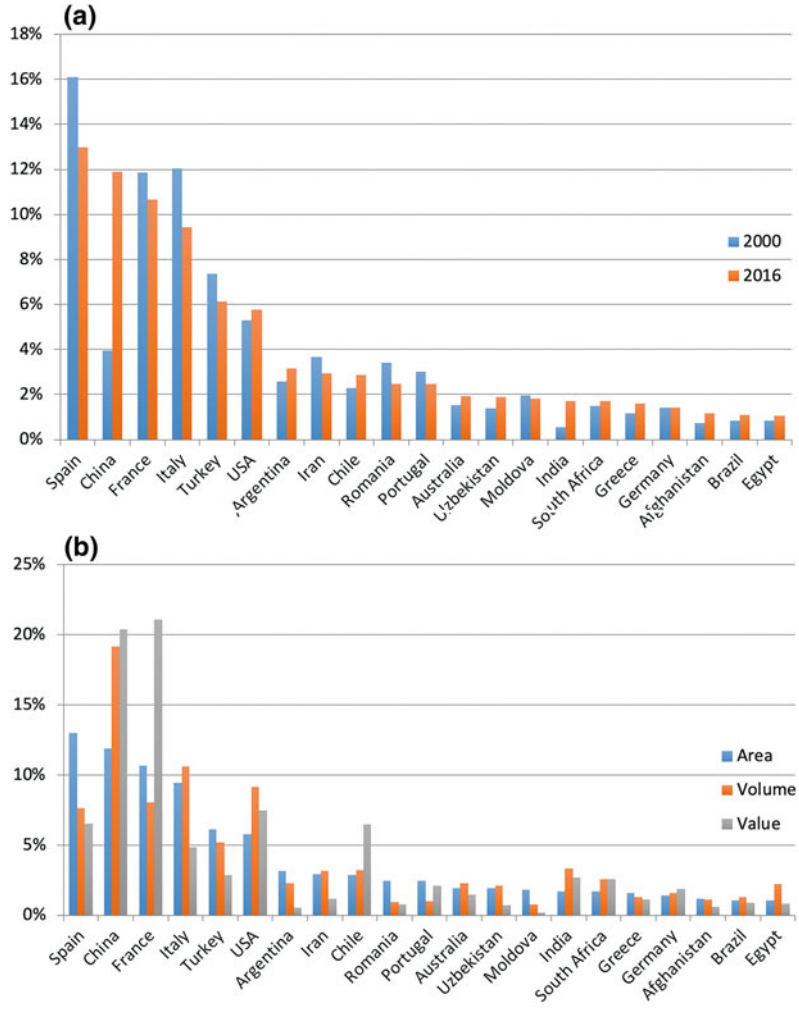

Fig. 1.1 Global distribution of grape area in 2000 and 2016, and area, production volume and value in 2016—top 20 countries by area in 2016. *Source* Created by the authors using data from FAOSTAT (2018). **a** National shares of global grape area, 2000 and 2016, %. **b** National shares of global grape area, production volume, and value, 2016, %

Total production, also, is concentrated among a few countries, but the ranking is slightly different reflecting differences in end uses and average yields. The top five countries in terms of quantity produced (now China, Italy, the USA, Spain, and France) accounted for 42.2 MT, more than half of the total of 77.4 MT, and just 15 countries accounted for 63.8 MT, more than four-fifths of the total. Country rankings change again when we look at value of production, reflecting differences in average unit values among countries, especially for wine grapes. In terms of value of production, the top five countries are France, China, the USA, Spain, and Chile.

These country rankings reflect both the historical origins of grape production in the Old World and the development of grape production in the New World, especially in recent decades. Whether in the New World or the Old World, grapes are grown in mid-latitude regions where temperatures during the growing season average 13–21 °C (Jones 2006), predominantly

Table 1.2 Production from top 20 grape-producing countries and world, 2000 and 2016

Country	2000		2016			Growth in production 2000–2016
	Production	Share of world total	Production	Share of world total	Cumulative share	
	KT	%	KT	%	%	%
China	3281.7	5.2	14,763.0	19.1	19.1	349.9
Italy	8869.5	14.0	8201.9	10.6	29.7	−7.5
USA	6973.8	11.0	7097.7	9.2	38.8	1.8
France	7762.6	12.2	6247.0	8.1	46.9	−19.5
Spain	6539.8	10.3	5934.2	7.7	54.6	−9.3
Turkey	3600.0	5.7	4000.0	5.2	59.7	11.1
India	1130.0	1.8	2590.0	3.3	63.1	129.2
Chile	1899.9	3.0	2473.6	3.2	66.3	30.2
Iran	2097.2	3.3	2450.0	3.2	69.4	16.8
South Africa	1454.7	2.3	2008.8	2.6	72.0	38.1
Australia	1311.4	2.1	1772.9	2.3	74.3	35.2
Argentina	2459.9	3.9	1758.4	2.3	76.6	−28.5
Egypt	1075.1	1.7	1716.8	2.2	78.8	59.7
Uzbekistan	624.2	1.0	1642.3	2.1	80.9	163.1
Germany	1361.0	2.1	1225.6	1.6	82.5	−10.0
Greece	667.6	1.1	990.3	1.3	83.8	48.3
Brazil	1024.5	1.6	984.5	1.3	85.0	−3.9
Afghanistan	330.0	0.5	874.5	1.1	86.2	165.0
Portugal	913.6	1.4	773.9	1.0	87.2	−15.3
Romania	1295.3	2.0	736.9	1.0	88.1	−43.1
Other	8881.0	14.0	9196.4	11.9	100.0	3.6
World	**63,552.7**	**100.0**	**77,438.9**	**100.0**		**21.8**

Source Created by the authors using data from FAOSTAT (2018)

in river valleys near the coast, often with a Mediterranean-type climate. Since growing season duration and temperatures have a major influence on grape ripening and fruit quality, within this broad landscape particular cultivars have been developed to be grown for particular end uses and in specific regions and sub-regions (see, e.g., Jones 2018).

The economic geography of grape production has been shifting over time, reflecting changes in both supply and demand for grape products among diverse countries. On the supply side, along with changes in technology of production and in the availability of labor and other inputs, changes in climate have begun to influence where particular cultivars can profitably be grown for particular end uses. On the demand side, along with changes in other sociodemographic factors, changes in income have implications for the mixture of grape products demanded given relatively high income elasticities of demand for premium wine versus basic wine, and for fresh versus dried grapes (see, e.g., Fuller and Alston 2012).