

A concise operating manual for

INTELLIGENT CROP BREEDING



Shoaib Ur Rehman
Jin-Ying Gou
Weilong Guo



Authors

Dr. Shoaib Ur Rehman
Institute of Plant Breeding and
Biotechnology, Muhammad Nawaz
Shareef University of Agriculture,
Multan, Pakistan

Dr. Jin-Ying Gou
Wheat Genetics and Genomics
Centre, China Agricultural
University, Beijing, China

Dr. Weilong Guo
Wheat Genetics and Genomics Centre,
China Agricultural University, Beijing,
China

ISBN: 978-969-7889-21-1

DOI: 000-000-00-00

© Global Science Publisher. 2026

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations used in critical reviews or other noncommercial uses permitted by copyright law.

The use of general descriptive names, trademarks, or service marks in this publication does not imply, even in the absence of a specific statement, that such names are exempt from relevant protective laws and regulations and therefore free for general use.

The publisher, authors, and editors have made every effort to ensure the accuracy of the information contained in this publication. However, they assume no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein. The publisher remains neutral with regard to jurisdictional claims in published materials and institutional affiliations.

This imprint is published by **Global Science Publisher**.
Registered address: Old Shujabad Road, Multan, Pakistan.

Preface

Agricultural productivity is one of the defining challenges at present. The demand for climate smart crops has never been greater in the face of climate change and population surge. This requires redesigning of agricultural science beyond conventional procedures and protocols to embrace a new era of precision agriculture.

Interdisciplinary convergence of technology, life and data science is the need to this era. Prevailing tools for gene sequencing permit us to explore blueprints of plant genetic background with high precision. This volume “A Concise Operating Manual for Intelligent Crop Breeding” is a result of a teamwork with international researchers who are enthusiastically shaping this new frontier. Our aim is to offer a comprehensive and practical guide that bridges the gap between academic theory and laboratory application. We anticipate that the effective application of this approach hinges on a solid base in basic laboratory practices. Without high-quality of molecular samples and careful preparation of molecular biology experimentation, even the state-of-the-art techniques will produce erratic outcomes.

We provide robust protocols for high quality isolation of nucleic acids which is a prerequisite of high throughput genotyping. More focus is given to Cleaved Amplified Polymorphic Sequences (CAPS), derived CAPS (dCAPS) and Kompetitive Allele Specific PCR (KASP) assays. The techniques are reliable to genotype breeding populations for SNP-based markers making the backbone of modern crop breeding.

“A Concise Operating Manual for Intelligent Crop Breeding” is hence designed as basic resource for a broad researcher- from postgraduates embarking on their careers to veteran molecular plant breeders and researchers looking to integrate these techniques into their work.

Dr. Shoaib Ur Rehman

Date: 30-June-2026

Table of Contents

Chapter 01	1
SNP Genotyping in Plant Breeding	1
1.1 Introduction to Genotyping	1
1.2 Genomic Location of SNPs	2
1.3 Applications of SNPs in Plant Breeding	4
1.4 SNP Detection Methods	5
References	6
Chapter 02	8
SNP-based Primer Designing	8
2.1 Guidelines for Primer Designing	8
2.1.1 Primer Length	8
2.1.2 Target Sequence Length	9
2.1.3 Primer Melting Temperature (T_m)	9
2.1.4 Primer Annealing Temperature (T_a)	9
2.1.5 GC Content and GC Clamp	10
2.1.6 Nucleotide Runs and Repeats	11
2.1.7 Cross-Homology and Paralogous Sequences	11
2.1.8 Primer Secondary Structures	11
2.1.9 Degenerated Positions and Mismatches	12
2.2 Genome Specific Primer Designing in Polyploid crops	12
2.3 Primer Designing for SNP Calling	13
2.3.1 Sequence Retrieval from the Genomic Database	15
2.3.2 Exploring Desired Polymorphic Site	15
2.3.3 Alignment of the Sequence	15
2.3.4 CAPS and dCAPS-based Primers	17
2.3.5 Designed Primer Attributes	17
2.3.6 KASP-based Primer Design	20
2.3.6.1 KASP Primer Designing on [C/T] Polymorphic Site	21

2.3.6.2 KASP Primer Designing in Case of 3bp InDel	21
References	23
Chapter 03	25
Genomic DNA Extraction from Plant Tissue	25
3.1 Prerequisites for DNA Extraction	25
3.1.1 Solutions Buffer	25
3.1.2 Sterilization of Laboratory Consumables:	30
3.2 Plant Tissue Sampling	31
3.3 Procedure of DNA extraction	32
3.3.1 CTAB DNA Extraction Protocol	33
3.4 Troubleshooting	34
3.5 DNA Quantification and Quality Assessment	36
3.6 Dilutions of Extracted DNA Samples	37
References	38
Chapter 04	39
Polymerase Chain Reaction	39
4.1 Introduction to Polymerase Chain Reaction	39
4.1.1 Initialization Step in PCR	39
4.1.2 Denaturation Step in PCR	40
4.1.3 Annealing Step in PCR	40
4.1.4 Extension/Elongation Step in PCR	40
4.1.5 Final Elongation Step in PCR	40
4.1.6 Final Hold Stage in Thermal Cycler	41
4.2 PCR Recipe	41
4.3 Conditions to Perform Standard PCR	42
4.4 Steps to Perform a Successful PCR	43
4.5 Gel Electrophoresis	47
4.5.1 Problems and solutions	49
References	51
Chapter 05	54
Gel-free Genotyping (KASP, PARMS, SNP Probe)	54

5.1 Kompetitive Allele Specific PCR (KASP)	54
5.2 Prerequisite to Perform the KASP Assay	54
5.2.1 Chemicals Required to Perform the KASP Assay	55
5.2.2 Primer Dilutions for KASP Assay	55
5.2.2.1 Preparation of Stock Solution	56
5.2.2.2 Preparation of Working Solution	57
5.2.3 DNA Quality and Quantity Requirement for KASP Assay	58
5.2.4 Plate Layout Preparation	58
5.2.5 Recipe to Perform the KASP Assay	58
5.2.6 Thermal Cycler Conditions	60
5.3 Penta-primer Amplification Refractory Mutation System (PARMS)	60
5.3.1 Primer Dilutions for PARMS Assay	60
5.3.2 Recipe to Perform the KASP Assay Using the PARMS Master-mix	61
5.4 Procedure to Perform the Gel-free Genotyping	62
5.4.1 PCR Amplification Using BIO-RAD CFX 96-well Real-time PCR	62
5.4.2 KASP Plate Read Using CLARIOstar ^{Plus} Microplate Reader	73
5.4.3 Data Generation by MARS Software	80
References	94
Chapter 06	100
BIGD Data-derived Markers	100
6.1 Public Database	100
6.1.1 NCBI Sequence Read Archive	100
6.1.2 Open Databases	101
6.2 Publicly Available Sequencing Data Website for Cereal Crops	103
6.2.1 Wheat Genome Sequence Database	103
6.2.2 Maize Genetics and Genomics Database	104
6.2.3 Rice SNP-Seek Database	105
6.3 Personalized and Customized Sequencing Data	105
6.3.1 Creating Personalized and Customized Sequencing Datasets	105
6.3.2 Processing Personalized Genome Sequence Data	106

6.4 Constructing the Genome Variation Website	107
6.5 Screening Differential Locus Between Different Genomes	108
6.6 Developing Molecular Markers	109
References	110
Chapter 07	111
Genotyping with High-throughput Sequencing Data and Applications	111
7.1 Introduction to Second Generation Sequencing Data	111
7.1.1 Development in NGS technology	111
7.1.2 NGS Principle	112
7.1.3 NGS Data Format	112
7.2 Mapping Reads to Reference Genome	113
7.2.1 Preprocessing	113
7.2.2 Indexing the Reference Genome	113
7.2.3 Alignment with Reference Genome	114
7.2.4 Post-Alignment Processing	114
7.3 Variation Calling	115
7.3.1 Preprocessing and Alignment with Reference Genome	115
7.3.2 Base Quality Score Recalibration (BQSR)	115
7.3.3 Variant Calling Using GATK	115
7.3.4 Joint Genotyping for Population Studies	115
7.3.5 Post-Processing and Filtering	116
7.4 Querying Variation with SnpHub	116
7.4.1 Viewing Basic Information of the Current Dataset	117
7.4.2 Variant Data Table Query (VarTable)	118
7.4.3 Heatmap Visualization of Variation Data	119
7.4.4 Haplotype Network Analysis of specific genes/intervals	119
7.4.5 Phylogenetic tree analysis of sample lists or sample sets in specified intervals	120
7.4.6 Visualizing the genotypes geographically	120
7.4.7 Displaying variation frequency and function on gene structure	121
7.4.8 Creating Consensus Sequence for an Individual	121

7.5 Haplotype Analysis	122
7.5.1 Definition of Haplotype	122
7.5.2 Main Components of Haplotype Analysis	122
7.5.3 Application Scenario	123
7.5.4 Common Tools for Haplotype Inference	124
7.6 Genome-Wide Association Study (GWAS)	124
7.6.1 Basic Principles of GWAS	124
7.6.2 Applications of GWAS in Crop Genetics and Breeding	125
7.6.3 Future Development Directions in GWAS	125
References	126

Chapter 01

SNP Genotyping in Plant Breeding

1.1 Introduction to Genotyping

Meeting the needs of the ever-increasing global population requires a broader understanding of the diversity and distribution of plant species. Extreme weather conditions and climate change have been identified as significant contributors to agricultural losses and crop deterioration (Backlund et al. 2008). Although significant improvements occurred during the last century in agricultural production (Bohra et al. 2014), further advancements are required to tackle new challenges. Today, agricultural fields still require new crop varieties due to ongoing cultural and social changes. Plant producers must meet market demands, customers preferences, and agronomic challenges. Significant improvements made so far have been attained by traditional breeding techniques. In contrast, biotechnology plays a crucial role in shaping future agricultural prospects and is a prerequisite for accelerating crop development (Lucht 2015). The research and practical application of genetic markers in agricultural biotechnology hold significant promise for future crop breeding (Lateef 2015). DNA markers associated with yield are commonly used in cultivar development for various crops, including maize (*Zea mays*) (Ribaut and Ragot 2007), rice (*Oryza sativa*) (Jena and Mackill 2008), tomato (*Lycopersicon esculentum*) (Tiwari et al. 2022), and wheat (*Triticum aestivum*) (Song et al. 2023).

A naturally occurring genetic variation at a single base site/position in a DNA sequence is referred to as a single nucleotide polymorphism (abbreviated as SNP). Generally, SNPs involve a change from one nucleotide base to another through transversions or transitions. However, single-base Insertion/Deletion (InDels) are also classified as SNPs. Two-nucleotide alterations and small InDels (spanning a few nucleotides) are frequently grouped with SNPs in genotyping pipelines. Plant scientists investigate whether and how SNPs influence phenotypic traits. These genetic variations at specific positions serves as effective molecular markers in both animal (Johnsson 2023) and plant species (Bohra et al. 2014; Ribaut and Ragot 2007; Song et al. 2023; Tiwari et al. 2022). The distribution pattern of these genetic variations differs across species, and even the genetic loci contributing to disease