Institut für Nutzpflanzenwissenschaften und Ressourcenschutz

-Pflanzenernährung-

Genetic mapping of quantitative trait loci associated with arsenic tolerance and accumulation in rice (*Oryza sativa* L.)

Dissertation

zur Erlangung des Grades

Doktor der Agrarwissenschaften

(Dr. agr.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Varunseelan Murugaiyan

aus

Coimbatore, India

Bonn 2019

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universtät Bonn

Referent: Prof. Dr. Michael Frei

Koreferent: Prof. Dr. Jens Léon

Tag der mündlichen Prüfung: 14.06.2019

This dissertation is submitted as a “Monograph” that covers five chapters resulting from two peer-reviewed (open access) publications that are listed as following:

Published papers:

1. Jauhar Ali †, Muhammad U †. Aslam; Rida Tariq †, **Varunseelan Murugaiyan** †, Patrick Schnable, Corinne M. Nazarea, Jose Hernandez, Muhammad Arif, Jianlong Xu, Zhikang Li. Exploiting the Genomic Diversity of Rice (*Oryza sativa L.*): SNP-typing in 11 Early-Backcross Introgression Breeding Populations**,** Front Plant Sci (2018) 9:1–10. (†Equal contribution)

2. **Varunseelan Murugaiyan,** Jauhar Ali, Anumalla Mahender**,** Muhammad U. Aslam, Jewel, Pang, Corinne M. Nazarea, Lin-Bo Wu, Michael Frei, and Zhikang Li. Mapping of genomic regions associated with arsenic toxicity tolerance in a backcross breeding population of rice (Oryza sativa L.), Rice (2019) 12: 61.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenine
ALT	Alternative allele
ANOVA	Analysis of variance
As	Arsenic
$As^{(III)}$	Arsenite
$As^{(V)}$	Arsenate
AWD	Alternate Wetting Drying
BC	Back cross
bp	Base pair
BRILs	Backcross Recombinant Inbred Lines
C	Cytosine
CDS	Coding region of sequence
DNA	Deoxyribonucleic acid
Dp	Donor parent
DSR	Direct Seeded Rice
F1	Filial generation one
G	Guanine
GBS	Genotyping By Sequencing
GF-AAS	Graphite Furnace Atomic Absorption Spectroscopy
GRLs	Genome Reduction Levels
GSR	Green Super Rice
GWAS	Genome Wide Association Mapping
HMA	Heavy metal associated protein
ILs	Introgression lines
K	Phosphorus
Kb	Kilo base pairs(1,000 bp)
LD	Linkage disequilibrium
LMD	Low-Missing Data
MAGIC	Multiparent advanced generation intercross
Mb	Mega base pairs(1,000,000 bp)
mg kg^{-1}	Milligram per kilogram

MLM	Mixed linear model
MRP	Multidrug resistance protein
N	Nitrogen
O.sativa	Oryza sativa
OPT	Oligopeptide transporter
P	Potassium
PCA	Principal Component Analysis
PHRED	DNA base calling score
ppm	Parts per million
QTL	Quantitative Trait Locus
REF	Reference allele
Rp	Recurrent parent
Si	Silicon
SNP	Single Nucleotide Polymorphism
SPAD	Soil Plant Analysis Development
T	Thymine
tGBS®	Tunable Genotyping By Sequencing
V-PPase	Vacuolar pyrophosphates
WTR1	WEED TOLERANT RICE-1

LIST OF FIGURES

Supplementary figure chapter 2

Supplementary figure chapter 3

LIST OF TABLES

SUMMARY

Rice (*Oryza sativa.* L) is one of the world's most important staple crops, consumed by more than half of the world's population, and it plays a significant role in the entry of mineral nutrients into the food chain. Arsenic (As) is an unwanted toxic mineral that threatens the major rice-growing regions in the world, especially in South Asia. Rice production in Bangladesh and India depends on As-contaminated groundwater sources for irrigating paddy fields and in the presence of elevated amounts of As in the topsoil. Arsenic accumulating in rice plants has a significant negative effect on human and animal health. Understanding the genetic basis of As tolerance and uptake in rice will provide useful information for the breeders to develop As safe varieties. This thesis is structured into four major parts: (I) screening diverse genotypes for a new source of As tolerance. Arsenic toxicity leads to significantly negative plant responses, including reduced germination ability, reduced biomass, stunted plant growth, reduced leaf chlorophyll content, and increased shoot As concentration. Under field condition, As accumulation in grain vary considerably with the genotype. In general, *Indica* subspecies were more tolerant than those from the *Japonica* subspecies. (II) Development of two permanent mapping populations using backcross breeding approach. Population one consist of 230 backcross lines derived from a cross between *Indica* × *Japonica* varieties and genotyped with 6K single nucleotide polymorphism (SNP) Bead chip. Population two consist of 563 lines derived from the cross between one common parent and eleven donor parents. Genotyped with tunable genotyping-by-sequencing (tGBS) approach resulting in 55,239 SNPs. (III) Population one was phenotyped under 10 ppm As toxicity stress for 18 days. In the second population, lines derived from the four crosses out of eleven crosses were considered for phenotyping under 5, 10 and 15 ppm As stress for 18 days and field screening was carried out for two consecutive seasons. As accumulation is shoot was directly proportional to As concentration in the treatment. In field conditions As content in the plant was considerably high in the dry season than the wet season. (IV) Quantitative trait locus (QTLs) mapping for As related traits was carried using association analysis. Marker-trait association was determined for As-related traits. QTLs were detected on several chromosomes (1, 2, 3, 5, 6, 8, 9, 11 and 12) indicating a complex genetic structure for As toxicity tolerance. Ten QTLs were mapped for relative tolerant traits, six QTLs for As content in shoots, two QTLs for As content in plant root and two QTLs for As content in grain. Several common loci were found among the tolerance trait and As content shoot traits. QTLs mapped on chromosome 8 and 6 were found to be consistent with a previous study and co-localized for As concentrations in grain at maturity and shoot phosphorus (P) concentrations at the seedling stage.

ZUSAMMENFASSUNG

Reis (*Oryza sativa.* L) ist eines der wichtigsten Grundnahrungsmittel der Welt, das von mehr als der Hälfte der Weltbevölkerung konsumiert wird und eine wichtige Rolle bei der Aufnahme von Mineral- und Nährstoffen in die Nahrungskette spielt. Die wichtigsten Reisanbaugebiete der Welt, insbesondere in Südasien, werden jedoch durch Arsen (As), einem unerwünschten toxischen Element, bedroht. Der Anbau von Reis in Bangladesch und Indien ist abhängig von As-kontaminierten Grundwasserquellen, die zur Bewässerung der Reisfelder eingesetzt werden, wodurch sich der As-Gehalt im Oberboden erhöht. Eine Anreicherung von As in Reispflanzen hat einen erheblich negativen Einfluss auf die Gesundheit von Mensch und Tier. Ein Verständnis der genetischen Grundlagen über die Toleranz von Reis gegenüber As sowie die Aufnahme von As in die Pflanze, wird nützliche Informationen für Züchter liefern, um As sichere Sorten zu entwickeln. Diese Arbeit ist in vier Hauptteile gegliedert: (I) Screening verschiedener Genotypen auf eine neue Quelle von As-Toleranz. Die Arsentoxizität führt zu signifikant negativen Reaktionen der Pflanzen, einschließlich reduzierter Keimfähigkeit, reduzierter Biomasse, verkümmertem Pflanzenwachstum, reduziertem Blattchlorophyllgehalt und einer erhöhten As-Konzentration im Spross. Unter Feldbedingungen, variiert die As-Anreicherung im Korn stark mit dem Genotyp. Im Allgemeinen waren die *Indica*-Unterarten toleranter als die der *Japonica*-Unterarten. (II) Entwicklung von zwei dauerhaft abgebildete Populationen unter Verwendung der Rückkreuzung. Die erste Population besteht aus 230 Rückkreuzungs-Linien, die von einer Kreuzung aus *Indica* × *Japonica*-Sorten abgeleitet und mit 6K Single Nucleotide Polymorphism (SNP) Beadchip genotypisiert sind. Die Population zwei besteht aus 563 Linien, die aus der Kreuzung zwischen einem gemeinsamen Elternteil und elf Spendereltern abgeleitet und mit einem abstimmbaren Genotypisierungs-by-sequencing (tGBS)-Ansatz genotypisiert wurden, was zu 55.239 SNPs führte. (III) Population eins wurde 18 Tage unter 10 ppm As-Toxizitätsstress phänotypisiert. In der zweiten Population wurden Linien, die von vier Kreuzungen aus den elf Kreuzungen abgeleitet wurden, für die Phänotypisierung unter 5, 10 und 15 ppm As-Stress für 18 Tage sowie als Feldscreening für zwei aufeinanderfolgende Saisons betrachtet. Die Anreicherung von As im Spross stand im direkten Zusammenhang zu der As-Konzentration der Behandlung. Unter Feldbedingungen war der As-Gehalt in der Pflanze während der Trockenzeit deutlich höher im Vergleich zur Regenzeit. (IV) Quantitative Trait Locus (QTLs) Mapping für die mit As verbundenen Traits wurden mittels Assoziationsanalyse durchgeführt. Die Marker-Trait-Zuordnung wurde für As-bezogene Merkmale bestimmt. QTLs wurden auf mehreren Chromosomen (1, 2, 3, 5, 6, 8, 9, 11 und 12) nachgewiesen, was auf eine komplexe genetische Struktur für die Toleranz gegenüber As-Toxizität hinweist. Zehn QTLs wurden für relativ tolerante Merkmale kartiert, sechs QTLs für den As-Gehalt im Spross, zwei QTLs für As-Gehalt im Pflanzenwurzel und zwei QTLs für As-Gehalt im Korn. Mehrere gemeinsame Loci wurden unter den Toleranzmerkmalen und dem As-Gehalt im Spross gefunden. QTLs, die auf den Chromosomen 8 und 6 kartiert waren, stimmten mit der vorherigen Studie überein und wurden für die As-Konzentrationen im Korn zum Reifestadium sowie für die Phosphor (P)-Konzentrationen im Keimlingsstadium, kolokalisiert.

1. INTRODUCTION

1.1. Rice production and global demand

Rice (*Oryza sativa.* L) commonly known as Asian rice is one of the world's most important staple crops, consumed by more than half of the world's population, and it plays a significant role in the entry of mineral nutrients into the food chain (Global Rice Science Partnership 2013; Yorobe et al. 2016). Currently, rice is planted in 166 million hectares worldwide, nurturing some four billion people, and is harvested annually with a total worth of USD 203 billion (Global Rice Science Partnership 2013). In Asia, rice is the primary cereal crop where 90% of the world's rice is produced and consumed (Bouman et al. 2013; Li and Ali 2017). About 480 million metric tons of milled rice is produced annually, China and India alone accounts for ~50% of the rice produced and consumed (Muthayya et al. 2014). About 23% of the total calories consumed by humans come through the consumption of rice (Ashikari and Ma 2015). Post Green Revolution, rice productivity has reached a very high level up to >8 t/ha in the irrigated areas of many countries (Li and Ali 2017). However, rice production and productivity need to keep pace with a growing global population estimated to reach 9.1 billion by 2050 (FAO 2009; Tyczewska et al. 2018). The predicament caused by climate change and a burgeoning population is leading to increased food insecurity and poverty. It is imperative to hasten the rate of genetic improvement efforts to meet the challenges of biophysical and socio-natural constraints. Improved breeding rice cultivars using cutting-edge biotechnological tools and efficient delivering within a shorter time frames is the fundamental solution to meet the challenges.

1.2. Origin and cultivation of rice

Advancement of human civilisation largely depends on the crop domestications and rice (*Oryza sativa* L.) belongs to the grass family (Poaceae) domestication is one of the most significant events in the history of agriculture evolution (Khush 2011; Molina et al. 2011; Huang et al. 2012). There are twenty-one wild species and two commonly cultivated species of genus *Oryza* (Ge et al. 2002). Asian cultivated rice *O. sativa* is domesticated from the wild ancestor *Oryza rufipogon* and is cultivated all over the world in an area ranging from latitude 53° to 40° (Brady 2012; Huang et al. 2012). The African cultivated rice *O. glaberrima* is domesticated from the wild ancestor *Oryza Barthii* is grown on a small scale in West Africa (Purugganan 2014; Wang et al. 2014).

(Purugganan 2014; Meyer et al. 2016). It is believed that Asian rice cultivation began simultaneously in multiple places in Asia over 6500 years ago (Gnanamanickam and Gnanamanickam 2009). Asian rice contains cultivars into two pivotal subspecies, *Indica* and *Japonica* (Eizenga et al. 2012). *Indica* subspecies were possibly domesticated in the foothills of the Himalayas in Eastern India and *Japonica* somewhere in the South China region. The *Indica* rice spread throughout the tropics and subtropics mostly closer to the equator (Svedberg et al. 2006). The *Japonica* rice moved northward from South China and became the temperate *Japonica*. They also moved southward to Southeast Asia and from there to West Africa and Brazil and became tropical *Japonica* (Khush 2011). Human induced selection and adaptation to diverse environments have resulted in numerous cultivars. It is estimated that about 1,20,000 varieties of rice exist in the world represents, an enormous gene pool for genetic improvement of rice cultivars (Svedberg et al. 2006; Khush 2011). African cultivated rice is believed to been domesticated 2,000–3,000 years ago in the inland delta of the upper Niger river by peoples living in the floodplains (Nayar 2010; Khush 2011). African rice often shows more tolerance to fluctuations in water depth, iron toxicity, infertile soils, severe climatic conditions and human neglect, and exhibits better resistance to various pests and diseases. However, the yield is considerably lower than Asian rice (Linares 2002).

1.3. Abiotic stresses in rice

Maintaining a favourable rice supply-demand balance in the future depends mostly on the exploitation of the production capacity of the rain-fed ecosystem in low-income Asian countries (Li and Ali 2017). About 60% of the agricultural land in Asia, rice is grown under rain-fed conditions (Rao et al. 2015). The rain-fed ecosystem is the dominant one in the low-income countries of Asia, where the demand for rice is projected to remain active (Li and Ali 2017). Rain-fed regions in India contribute substantially toward 44% of rice production (Rao et al. 2015). Rice plants are subtle to various abiotic stresses. Abiotic stress is the primary factor undesirably affecting crop growth and yield worldwide (Gao et al. 2007). Various abiotic stresses limit rice production in rain-fed environments, which comprise about 35% of the global rice area (Li and Ali 2017). Critical abiotic stresses including extreme temperature, drought, submergence, salinity, iron toxicity and deficiencies of phosphorus (P) and zinc (Zn)

are known to cause serious loss in terms of the yield (Lafitte et al. 2004; Wu et al. 2017; Li and Ali 2017).

Drought is a major constraint affecting rice production, especially in rain-fed areas across Asia (Babu 2010). Drought stress triggers various morphological, physiologically and biochemical process in rice. Reduced plant height, plant biomass, number of tillers, photosynthesis, transpiration, stomatal conductance, relative water content, chlorophyll content and abscisic acid content influences plant yield (Wang et al. 2011; Pandey and Shukla 2015). Traditional breeding programs aimed towards improved drought stress tolerance have had some success, but are limited by the multigenic nature of the trait. Salt-water intrusion in the coastal regions is another significant abiotic stress that limits rice production. Rice plant cannot tolerate excess salt in the soil, often salinity stress is compounded by nutritional deficiencies, mineral toxicities and drought stress (Gregorio et al. 2002). Thus, breeding rice for saline environments is challenging and multiple stress tolerance traits must be considered for crop improvement in the saline-prone areas (Gregorio et al. 2002; Pang et al. 2017). Ferrous toxicity is a major nutritional disorder that occurs in the flooded soils. It can affect the yield of rice in rain-fed production systems. Ferrous toxicity is associated with high concentrations of reduced iron ($Fe^{(2+)}$) in the soil solution. Although the first interface of the ($Fe^{(2+)}$) is in the roots, excessive uptake observed in several rice tissues resulting in considerable yield loss (Finatto et al. 2015; Frei et al. 2016).

The advances in physiology, genetics, and molecular biology have greatly improved our understanding of plant responses to stresses. For instance, incorporation of submergence tolerance 1 (SUB1) quantitative trait loci (QTL) identified in chromosome 9 from *Indica* cultivar FR13A makes rice plants persist up to two weeks of complete submergence (Xu et al. 2006; Perata and Voesenek 2007; Fukao and Bailey-Serres 2008). Using the marker-assisted breeding technique, SUB1 QTL was successfully transferred into mega varieties of Asia, expected to provide protection against increasing floods and increase crop security for the resource-poor farmers (Xu et al. 2006). Likewise, Pup1 (Phosphorus uptake 1) QTL identified in chromosome 12 from tolerant *Indica* landrace kasalath, allows plants to grow under low phosphorus (P) soil. Incorporation of Pup1 QTL in modern rice varieties makes a more cost-effective solution than relying on expensive fertiliser application (Ni et al. 1998; Wissuwa et al. 2002).

1.4. Arsenic contamination in rice-growing regions of Asia

Arsenic (As), is ubiquitously found in various inorganic and organic forms. Arsenic is a carcinogenic metalloid can enter into the environment by both natural and anthropogenic activities (Sturchio et al. 2013). Groundwater As contamination and its health effects in South-East Asian countries came to limelight in the year 1984 (Garai et al. 1984). A substantial part of the Ganga-Meghna-Brahmaputra plain with an area 569,749 km^2 and population over 500 million was at risk of As poisoning (Roychowdhury 2008). The amount of As that is permissible in drinking water is 10 $\mu g\ l^{-1}$ (Muhammad et al. 2010; Puri and Kumar 2012). However, very high levels of As contamination of groundwater and its adverse impact on human health have been reported in many countries of the world (Nriagu et al. 2007). The magnitude of this problem is quite severe in Bangladesh followed by West Bengal, India and China (Nickson et al. 2000; Mohan and Pittman 2007; Tareq et al. 2010; Mazumder 2013). Evidence has emerged in recent years that As-contaminated groundwater occurs commonly in other Asian countries including Cambodia, Myanmar, Pakistan, Nepal, Vietnam and the Kurdistan province of Iran (Figure 1.1) (Smith et al. 2000; Dipankar et al. 2004; Brammer and Ravenscroft 2009; Jiang et al. 2013). In the 1970s the use of surface water was abandoned mainly in the Bengal delta in response to severe health effects caused by microbial pathogens (Caldwell et al. 2003). This resulted in extensive usage of As-contaminated groundwater unexpectedly. The high-As groundwater is produced from shallow (<100 m) depths by domestic and irrigation wells in the Bengal Basin aquifer system (Sultana 2013). It has been reported that groundwater from shallow tube-wells (12–33 m) contains very high amounts of As. Alternatively, the water from deep tube wells (200–300 m) contains a lesser amount of As (<50 $\mu g\ l^{-1}$) (Hossain 2006). The subsurface mobilisation of As is mainly caused by a combination of chemical, physical and microbial factors and various theories have been propounded to explain the mechanism of As mobilisation (Anawar et al. 2003; Amini et al. 2008). Of these, the essential theories are the pyrite oxidation and oxy-hydroxide reduction (Hossain 2006) and the arsenic dissolution and released in the deltaic region have been modelled upon the contribution of microbes, organic matter as well as palaeosol formation (Gorny et al. 2015). Flooding induces reducing (anaerobic) conditions in soils (Pezeshki and DeLaune 2012) hence $As^{(V)}$ is reduced to $As^{(III)}$ and adsorbed $As^{(V)}$ is released as $As^{(III)}$. Alluvial and deltaic environments are

mainly characterised by reducing conditions which cause high As release in groundwater (Abernathy et al. 2001; Lee et al. 2008). When As enters the food chain, it causes widespread distribution throughout the plant and animal kingdoms (Mandal and Suzuki 2002). Both long and short term exposures are hazardous and can lead to skin, bladder, lung and prostate cancers, cardiovascular diseases, diabetes, anaemia as well as reproductive, developmental, immunological and neurological effects (Roy and Saha 2002; Ng 2005; Guha Mazumder 2008).

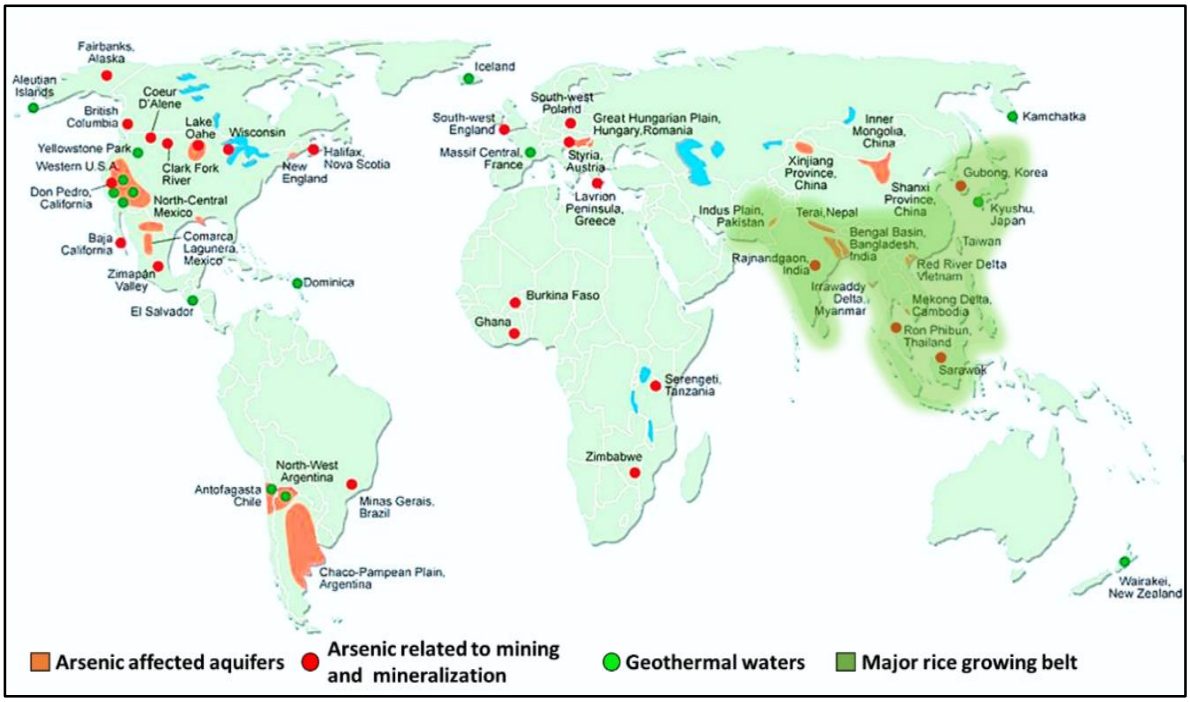


Figure 1.1. Worldwide distribution of arsenic contaminated regions and showing contaminated regions overlapping with major rice growing belt.
Image modified from British Geological Survey (Kinniburgh and Smedley 2001).

1.5. Arsenic contamination in rice

Arsenic (As) poisoning affects millions of people worldwide. Contamination of the agricultural soil with As from both anthropogenic and natural sources has occurred in many parts of the world and its accumulation in food crops may pose a severe health risk to humans (Nriagu et al. 2007). In most parts of the world, low levels of As are naturally present in agricultural soil, the background levels of As in agricultural soil are around 5 - 10 mg kg^{-1} and substantial variation in concentration was observed depending on the origin of the agricultural soil (Reimann and Garrett 2005; Hughes et al. 2011; Su et al. 2014). Chronic As exposure currently affects more than 150 million

people worldwide and the consumption of rice is one of the primary routes of As exposure (Bastías and Beldarrain 2016). About 30% of the total As ingestion was contributed through rice and rice products (Li et al. 2011; Talukder et al. 2011). Existence of As in the paddy soil is complicated, existing in both inorganic and organic forms, and differ distinctly under flooded (anaerobic) and non-flooded (aerobic) conditions (Meharg and Hartley-Whitaker 2002). The toxicity of As is associated with reduced soil condition (flooded soils), which increases the bioavailability of inorganic As and uptake into the rice (Zhao et al. 2010a; Islam et al. 2016). Generally, rice grown in flooded soil under reduced condition is more efficient in assimilating inorganic As into its grain, and the accumulation may adversely affect the nutritional quality of grain (Islam et al. 2016). Of the total As present in the rice grain, inorganic As alone constitutes for 54 %. Also, grain As content increases with increased As concentration in the paddy soil (Suriyagoda et al. 2018a). As content in rice plant decrease in the order of roots > leaves > grains, and grains: hull > bran polish > brown rice > raw rice> polished rice > cooked rice (Suriyagoda et al. 2018a).

1.6. Source of Arsenic in agricultural paddy fields

Traditionally, rice is grown in flooded paddy fields, which can harbour unwanted toxic heavy metals such as arsenic, cadmium, mercury, and lead, that represents a threat to human and cattle health (Norton et al. 2010; Pandey et al. 2015). Compared with the other major cereals, rice takes up larger amount of As, a highly toxic class I carcinogenic metalloid, which naturally occurs worldwide in the environment and is widely distributed in Earth's crust (WHO 2010). Long-term use of As-contaminated groundwater for irrigating crops has resulted in a significant increase in As in the topsoil, thereby contaminating the food chain (Pandey et al. 2015). In rice paddies, As is mobilised and transported to rice grains, where it can impact human health upon ingestion. Also, arsenic uptake can decrease rice yields and contribute to food insecurity (Banerjee et al. 2013). Rivers originating from the greater Himalayas carry As from their rock sediments to the densely populated rice-growing plains of South and Southeast Asia, making the primary rice-growing belt of Asia vulnerable to As contamination (Sohn 2014; Zhou et al. 2017). Also, climate change is threatening rice production in these areas. With frequent incidences of drought and saltwater intrusion in the Ganges-Brahmaputra deltas of India and Bangladesh, farmers depend on groundwater for irrigation of their crops (Guan et al. 2010; Marcaida et al. 2014; Wang

et al. 2015). This groundwater is an additional source of As discharged from the naturally As-rich aquifers (Nickson et al. 1998). More than 70 countries have reported that their groundwater is contaminated with As (Figure 1.1), placing 150 million people worldwide at risk of As exposure (Rahman et al. 2009; O. Akinbil and Haque 2012). Rice production in the major regions of Bangladesh, India and Vietnam uses As-contaminated groundwater for irrigation. Typically, 4 to 8 mg kg^{-1} of As occurs in flooded paddy soil, but the concentration can reach 83 mg kg^{-1} in some parts of Bangladesh and West Bengal regions of India (Zhang et al. 2007; Norton et al. 2010). The amount of As in the agriculture soil can be varied, due to the geographical location, environmental influences and irrigation season (Awasthi et al., 2017; Mwale et al. 2018). However, contamination of Arsenic through groundwater is found all over the world, and the level is comparatively high in river basin deltas of the south and Southeast Asia, the primary rice production belt (Panaullah et al. 2008; Banerjee et al. 2013).

1.7. Inorganic Arsenic species and their occurrence in the rice ecosystem.

Arsenic toxicity in rice plants triggers various symptoms that includs reduced germination rate, poor seed establishment, reduced photosynthetic rates, stunted plant growth, low biomass production, sterility-related yield loss, and a physiological disorder refered as straighthead disease. These symptoms are often confounded with other soil-related problems associated with rice (Abedin et al. 2002b; Rahman et al. 2008; Zhu et al. 2008). The toxic effects of As in rice depends upon their chemical form, with the inorganic As species being more toxic than the organic species and both form of As present in the soil (Abedin et al. 2002c; Tripathi et al. 2013). The most common inorganic species occur in the rice ecosystem are arsenate (As^{V}) and arsenite (As^{III}) (Figure 1.2), while the most common organic species are monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA). The toxicity of arsenic species is in the following order $As^{III} > As^{V} >$ MMA > DMA (Vahter and Concha 2010). The inorganic species of As found in the rice-growing environment are arsenite (As^{III}) and arsenate (As^{V}). In anaerobic flooded soil conditions such as submerged paddy fields, the reduced form $As^{(III)}$ dominates and in aerobic soil conditions such as upland paddy fields its oxidised form $As^{(V)}$ dominates (Figure 1.2) (Tripathi et al. 2013; Pandey et al. 2015). Both forms are toxic to rice plants (Batista et al. 2011).

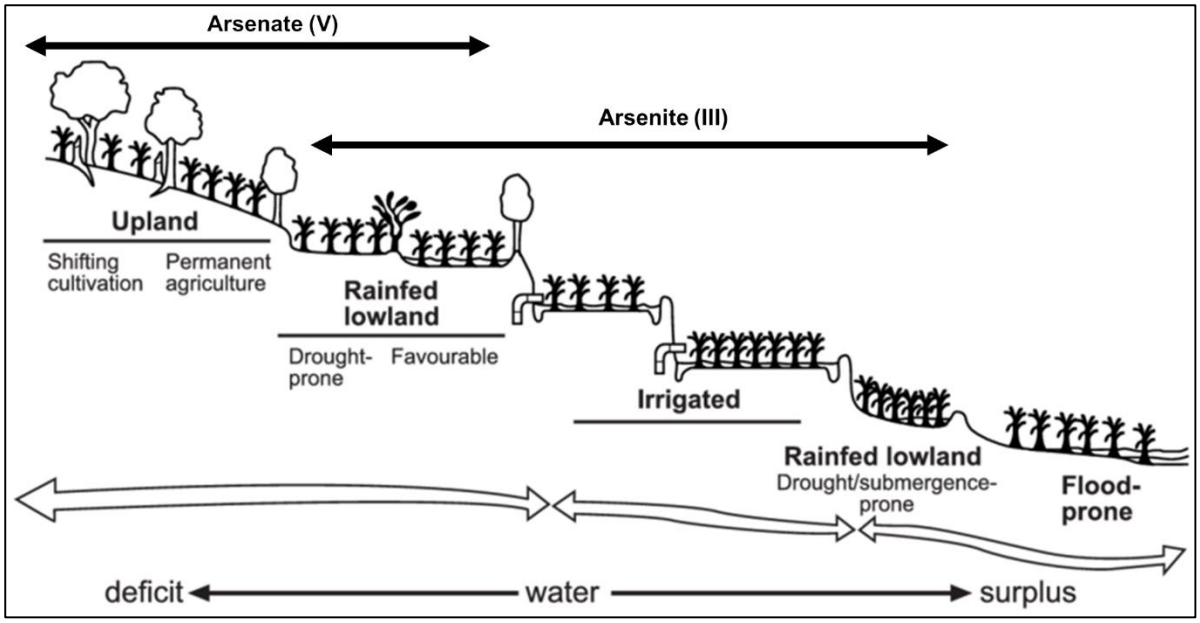


Figure 1.2. Inorganic Arsenic species found in the rice-growing environment (adapted and modified from Rice almanac 3rd edition(Maclean, J.L. 2013)).

1.8. Quantitative trait loci for Arsenic stress in rice

Genetic improvement and selection of rice cultivars with a lower concentration of As in the grain are essential to the develop As-safe varieties (Norton et al. 2012; Duan et al. 2017). Mapping QTLs to identify causative genes in populations is essential for trait improvement in breeding programs (Würschum 2012). Selection of appropriate of rice cultivars and genetic mapping of a chromosomal region of QTLs with lower concentration As in grain and tolerance to against As, which might be the practical methodology for reducing As contamination in rice and also to ensure superior grain yield level and quality in As-contaminated regions (Fig. 3). There are limited genetic linkage maps and QTLs information in rice genome for As stress tolerance traits and low As concentration in grains (Figure 1.3). Using backcross breeding populations has proven to be an effective strategy to dissect the genetic factors that underly the phenotypic complexity of nutrient-related traits with a simultaneous focus on varietal development (Xue et al. 2006; Ali et al. 2013; Wu et al. 2014). Although As interaction with rice has been well documented over the past two decades, only a limited number of QTL studies and genes associated with As have been reported for the development of As-safe varieties (Dasgupta et al. 2004; Zhang et al. 2008, 2014; Kuramata et al. 2013). *AsTol* on chromosome 6 was the first QTL reported in rice for root tolerance to $As^{(V)}$ and it is located close to a phosphate uptake QTL (Dasgupta et al. 2004). For

$As^{(III)}$, the first QTLs reported in rice are located on chromosomes 2 and 3 for As content in shoot and root, respectively, and chromosome 6 and 8 for As content in brown rice (Zhang et al. 2008). However, QTLs mapped for As tolerance in rice is not characterised to use in breeding programs for the development of As-safe varieties.

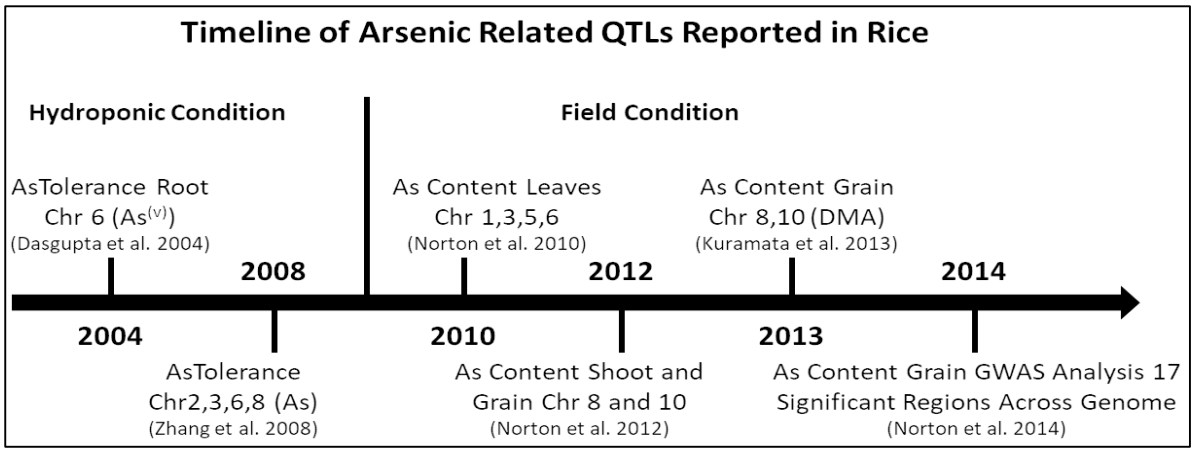


Figure 1.3. Timeline of Arsenic-related QTLs reported in rice

1.9. Genes associated with Arsenic transport in rice

Rice plants do not possess the naturally evolved As transporters (Pandey et al. 2015), instead, As competes with chemically similar essential minerals to enter the plant system (Wenzel and Alloway 2013). $As^{(III)}$ is physiochemically similar to the essential mineral silicon (Si) and thus competes with the Si uptake pathway. Alternatively, $A^{(V)}$ is physiochemically similar to essential mineral phosphorous (P) and uses P acquisition pathways to enter the root system, and for efflux toward the xylem and various tissues (Clemens 2006; Zhao et al. 2009; Yang et al. 2018). Most rice genotypes possess a mechanism to retain much of the toxic As burden in the roots. However, a genotype-dependent proportion of As is translocated into the shoots and other tissues, including grains of the rice plant (Carey et al. 2010; Pandey et al. 2015). Since 35-55% of rice is produced in irrigated conditions (Ali and Jewel 2018), $As^{(III)}$ contributes the major As species loaded into rice plants (Zhao et al. 2010b). $As^{(III)}$, which exists as the neutral molecule $As(OH)_3$, enters rice root cells through nodulin 26-like intrinsic proteins (NIPs), belonging to the aquaporin family of major intrinsic proteins (MIPs), which are non-permeable to $As^{(V)}$ (Ma et al. 2008; Zhao et al. 2009). Widely spread NIP aquaporin's also mediate the transport of a range of neutral molecules, including ammonia, urea, boric acid, and silicic acid (Zhao et al. 2010a). The silicon transporter in rice root *OsLSI1* has been suggested as the primary $As^{(III)}$

uptake protein, while $As^{(III)}$ efflux from rice root cells to the xylem takes place through *OsLSI2* silicon-mediated transporter (Figure 1.4) (Ma et al. 2008; Zhao et al. 2010b). However, none of the above genes associated with the $As^{(III)}$ and $As^{(V)}$ tolerance-low uptake genes were functionally characterised in rice to use in breeding programs for the development of As-safe varieties.

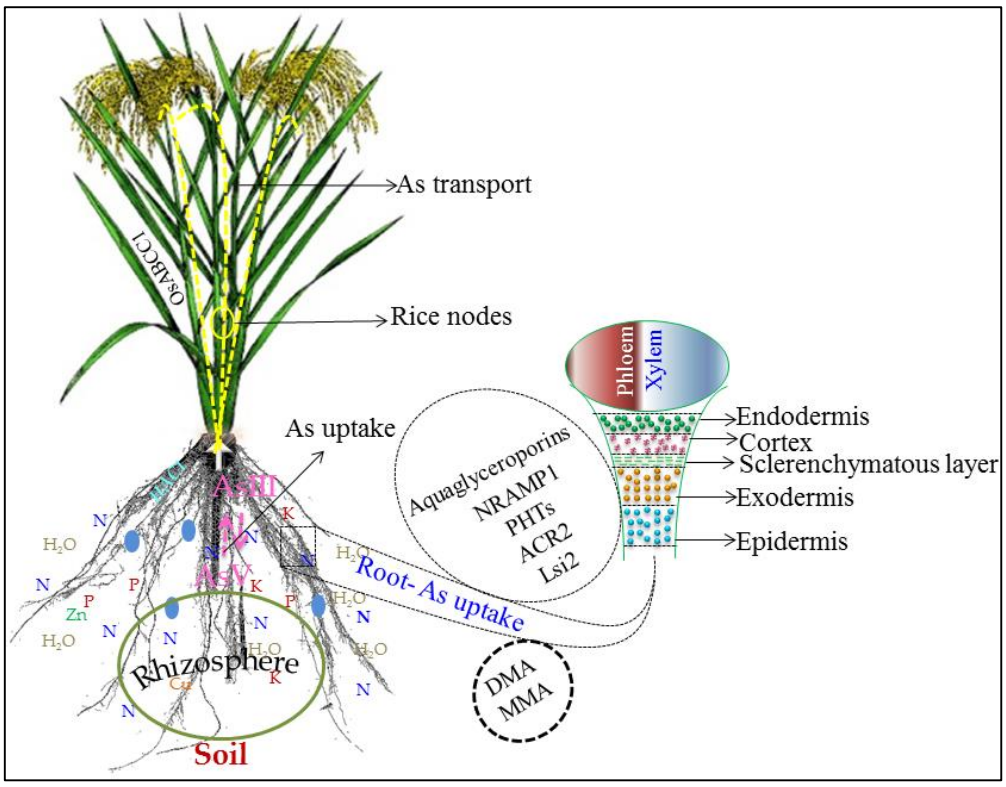


Figure 1.4. Schematic representation of genes involved in arsenic (As) transport in rice.

Rice nodes play a vital role in the As storage, distribution and serving as a filter restricting As transport from the root to shoot. As sequestrated into vacuoles is mediated by the *OsABCC1* (C-type ABC transporters (ABCC) during the reproductive stage. Dashed arrows and circles represent the different exogenous and endogenous genes and transcription factors influences to reduce the excess amount of As uptake and further accumulation in rice grain. In soil, the rhizospheric processes as the releasing oxygen from the roots, formation of iron plaque, and microbial oxidation mechanism, all of them are contributing to the $As^{(III)}$ oxidation to $As^{(V)}$ in soils and followed with the help of *HAC*1 (High Arsenic Content 1) arsenate reductases, $As^{(III)}$ influx and efflux transporters (*Lsi*1, *Lsi*2), aquaporins, phosphate transporters (Pht), and *NRAMP*1 plays a role in the As uptake and transport from root xylem and translocation, respectively.

1.10. Breeding rice for Arsenic contamination ecosystem

Governments of developing Asian nations are encouraging farmers to adopt modern technologies such as direct-seeded rice (DSR) and alternate wetting-drying (AWD) as a cost-efficient sustainable method to reduce the environmental footprint of rice production (Richards and Sander 2014; Singh Chauhan et al. 2015). Toxic levels of As in the topsoil can potentially affect the performance of DSR by hindering the most delicate rice germination process and early seed establishment of the rice growth cycle (Abedin and Meharg 2002; Shri et al. 2009). Water management using AWD was proposed as one of the strategies to control As bioavailability in the soil-plant system (Mitra et al. 2017). However, rice varieties suitable for the AWD system under As-toxic soil have not been developed and tested in field conditions for implementation (Suriyagoda et al. 2018). There is a substantial genetic variation among the rice genotypes in grain-As accumulation (Islam et al. 2016). Accumulation of As in rice grain involves several stages and borders which control the translocation of As from roots to grain, giving rise to potential differences among genotypes (Meharg and Zhao 2012). By identifying and breeding tolerant rice cultivars that restrict As translocation from root to shoot will be the sustainable solution for the As accumulation in rice.

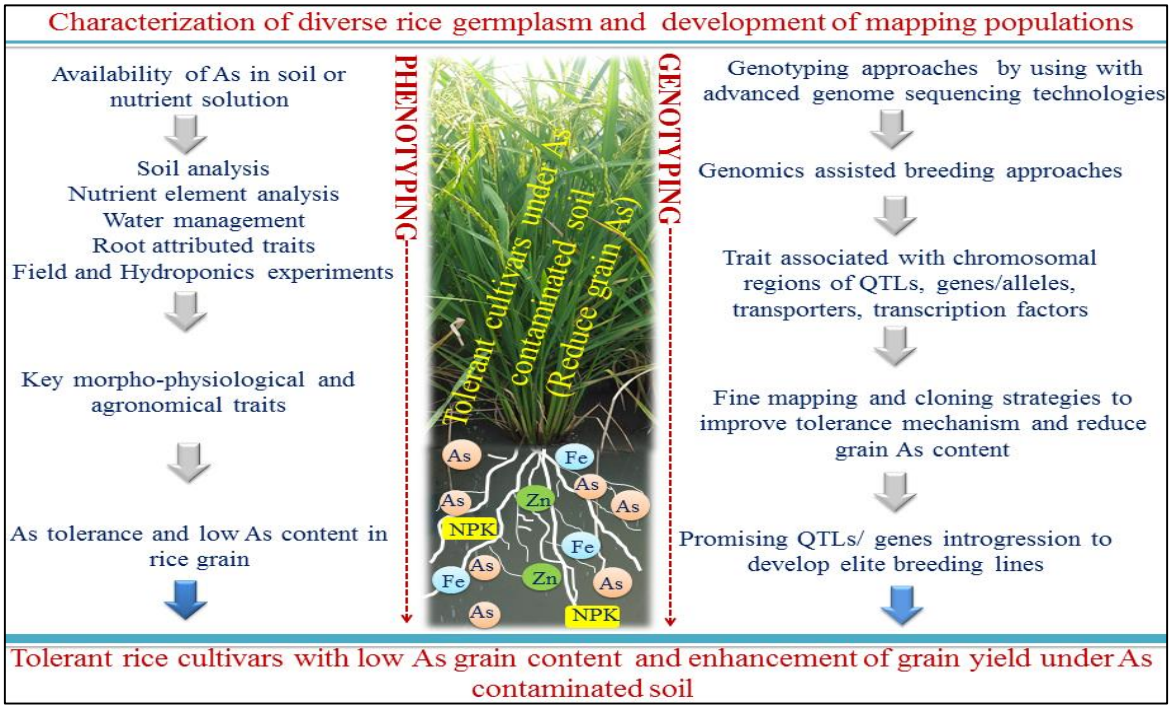


Figure 1.5. Integrated approaches of phenotypic and genomics-assisted breeding technologies for the development of As toxicity tolerant rice varieties.

1.11. Project aims

The overall objective of this study was to gain a better understanding of the genetics behind the arsenic toxicity stress in rice. While the toxicity caused by direct expose of As on humans and cattle health was well established, there were limited literature available regarding the consequence of the regular small concentration of As exposure through consuming As contain rice. Emerging reports suggest that rice, one of the staple food crops for billions of people, absorbs As from irrigation water and soil. Long-term ingestion of As through rice may increase the risk of various health problems, including cancer, heart disease, type 2 diabetes and decreased intelligence. High level of As exposure to the rice plants also induces toxicity symptoms, like poor plant establishment, yield loss due to sterility and stunted growth. Often these symptoms are mistaken for other soil-related problems. Therefore, it is necessary to reveal the genetic basis of As tolerance and accumulation in rice, which will help rice breeders to develop As-tolerant varieties with low As content in the grain.

The key aims of the project were to:

- Establishment and pre-screening of diverse genotypes for identification of arsenic tolerance and low accumulating genotypes.
- Development of permanent backcross breeding population from the identified genotypes for mapping genomic regions associated with arsenic stress.
- To conduct phenotyping-screening experiment under hydroponic and field condition for arsenic stress in the developed mapping population
- Development and application of high throughput genotyping methods in developed population for mapping genomic regions associated with arsenic stress.
- Mapping genomic regions associated with arsenic-related traits using association analysis.
- Identification of candidate genes in the mapped QTL region for functional characterisation in the feature.

These objectives were addressed with different approaches, including numerous plant breeding approach for population development, High throughput genotyping using Ion Proton System, phenotyping using various approaches, sensitive graphite furnace analysis of the sample for determining total As content and using various gene mapping methods to conduct mapping analysis.

2. RESULTS

2.1. Effect of Arsenic on seed germination

A pre-screening experiment was carried out to observe the effect of As on germinating seeds. Fifty-three genotypes used in the pre-screening experiment was carefully chosen to contain genotypes widely cultivated in the As-contaminated regions. Arsenic exposure showed significant inhibition in the germination ability among the studied genotypes. Germination percentages of different rice genotypes had different responses to As stress and germination decreased significantly with increasing As concentration in the treatment (Figure 2.2). The mean (average value of 53 genotypes) germination ability over control declined significantly with increasing As concentration in the treatment (p<0.01). The highest mean of (71.1%) germination was recorded in 5 ppm As treatment and the lowest mean of (19%) in 20 ppm As treatment. Tolerance percentage among genotypes in 5 ppm As treatment ranged between 29.5% in genotype X 21 to 104.5% in Haoannong. In 10 ppm As treatment, genotype kelmekri 77/5 exhibited the most reduced tolerance of <1% germination and genotype Haoannong showed high tolerance of 101%. In 15 ppm As treatment, *Indica* genotype BR11 from Bangladesh failed to germinate even after ten days of the incubation period. However, *Japonica* genotype Haoannong showed exceptional tolerance of 82.5% in 15 ppm As treatment. In 20 ppm As treatment, most of the evaluated genotypes tolerance percentage was reduced below <20%. However, genotype OM997 from Vietnam (*Aus* sub-species of cultivated rice *O.sativa*) showed significant tolerance of 68.09% in 20 ppm As stress (Supplementary data 1).

2.2. Genotype classification based on germination tolerance percentage

Fifty-three genotypes were grouped into four clusters established on the tolerance percentage under varying As treatment. Genotypes showing high tolerance >80% to As stress were clustered into highly tolerant genotypes and genotypes exhibiting poor tolerance percentage <50% across all the As stress treatment were grouped as highly susceptible genotypes. Genotypes were exhibiting only tolerance in lower concentration As treatment (5 and 10 ppm) as moderately tolerant genotypes and genotypes exhibiting susceptible from 10 ppm As treatment onwards as moderately susceptible (Figure 2.1). Except for genotype Haoannong, all the genotype form the *Japonica* sub-species of cultivated rice *O. sativa* were highly susceptible under As

stress. High yielding mega variety IR64 was moderately susceptible for germination under As toxicity stress. Genotype Bg300, BG304 from Sri Lanka showed a higher level of tolerance under As stress. Bangladesh origin genotypes BR11 and BR22 were highly susceptible for germination under As stress.

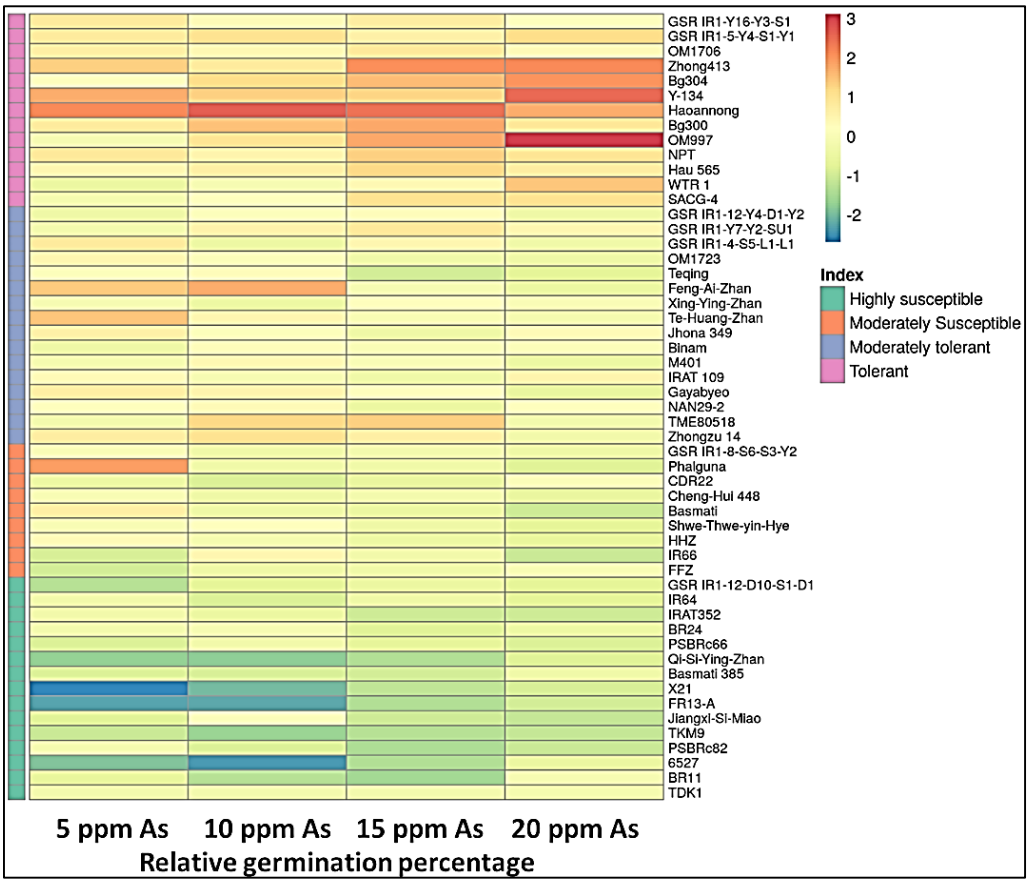


Figure 2.1. Grouping of fifty-three different cultivars based on germination tolerance under different arsenic stress.

As- Arsenic, ppm – parts per million

2.3. Effect of Arsenic on seedling growth parameters.

Effects of different concentrations of As (0, 5 and 15 ppm) on different morphological traits and As accumulation in young seedlings, were investigated in order to elucidate the toxicity of As in the early rice seedling stage. The means of different growth parameters, chlorophyll content, seedling height, root length, biomass accumulation were significantly influenced by the As concentration in the treatment for 18 days (Table 2.1 and Figure 2.2). In chlorophyll content, plant height, and root length showed significant variation between control and 15 ppm As treatment. Root and shoot, biomass accumulation was significantly affected with an increase in the treatment As

concentration. There was a significant difference between the control and treatment and the genotypes respond differently to the As treatment.

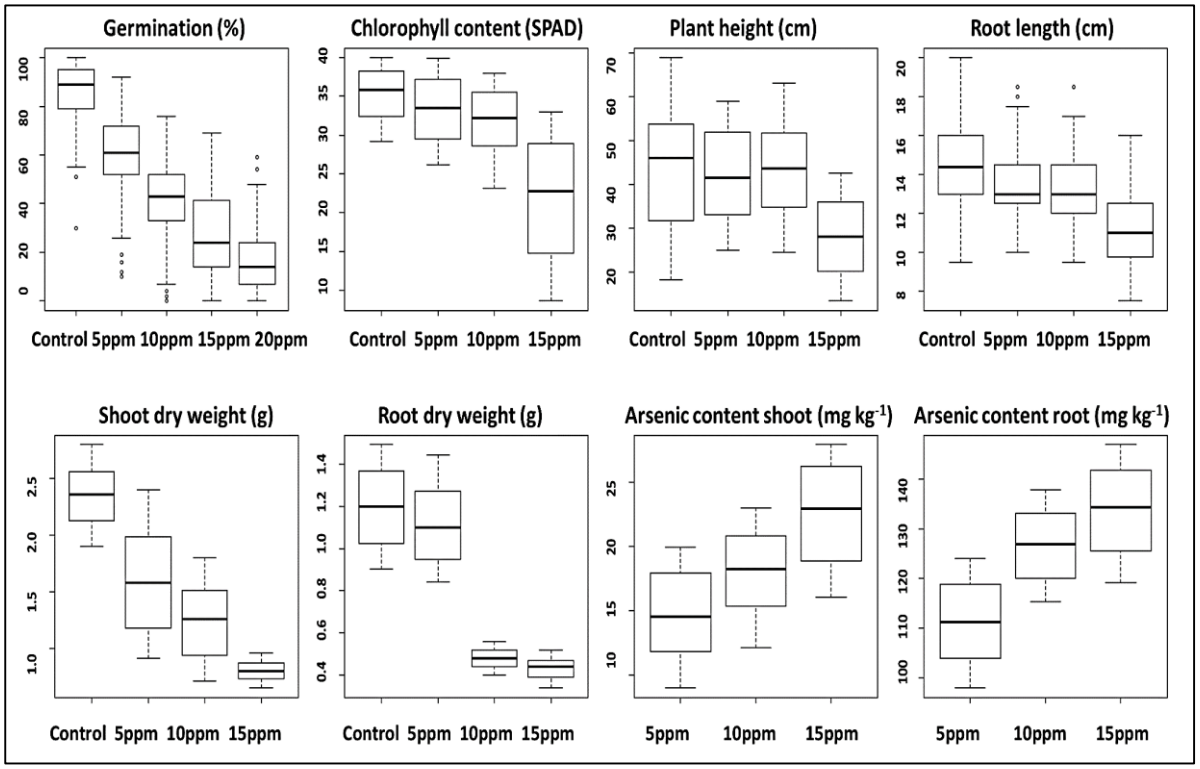


Figure 2.2. Box plot showing the distribution of phenotype values in fifty-three different cultivars under arsenic stress.

The thick line in the middle of the box is the median of the distribution while the lower and upper boundaries represent first and third quartile, respectively. Lower and upper whiskers are calculated based on 1.5 times the inter-quartile range. Control- without arsenic treatment, (5, 10, 15 and 20 ppm) represents different levels of arsenic treatments (n=3).

2.3.1. Arsenic accumulation in young seedlings

Accumulation of As in the shoot and root tissue increased significantly with the increase of As in the treatment (Figure 2.2). There were significant genotype differences in As accumulation in young seedlings after 18 days As treatment. In, 5 ppm As treatment, As content in shoot ranged from 9.25 mg kg^{-1} in Binam to 19.84 mg kg^{-1} in CDR22 genotype and root As content ranged from 98.25 mg kg^{-1} in Haoannong to 123.83 Cheng-Hui 448 genotype. In the 10 ppm As treatment, As content in shoot ranged from 12.37 mg kg^{-1} in GSR IR1-5-Y4-S1-Y1 to 22.95 mg kg^{-1} in OM1706 genotype and root tissue from 115.77 mg kg^{-1} in X21 to 137.65 in BR11 genotype. In the 15 ppm As treatment, As content in genotypes ranged from 16.08 mg kg^{-1} (Binam)

to 27.85 mg kg^{-1} (Xing-Ying-Zhan) and 119.86 mg kg^{-1} (Huang-Hua-Zhan) to 146.54 (BR11) in the shoot and root respectively. The overall results indicated that adequate genetic variability is present among the studied genotypes for As accumulation in young seedlings to As stress. The concentrations of As in seedlings of 53 varieties are presented in (Supplementary data 2).

Table 2.1: Descriptive statistics and ANOVA results for measured phenotypes.

	Trait	Chlorophyll content (SPAD)	Plant height (cm)	Root length (cm)	Shoot dry weight (g)	Root dry weight (g)	Arsenic content of shoots (mg kg^{-1})	Arsenic content of roots (mg kg^{-1})
Control	Min	29.43	20.36	9.50	1.91	0.94	ND	ND
	Max	39.91	68.23	18.83	2.80	1.43	ND	ND
	Mean	35.31	44.75	14.42	2.35	1.19	ND	ND
Treatment 5 ppm As	Min	26.30	25.67	10.17	0.92	0.85	9.25	98.25
	Max	39.84	58.72	17.83	2.39	1.44	19.84	123.83
	Mean	33.26	41.87	13.49	1.60	1.12	14.79	111.40
Treatment 10 ppm As	Min	23.31	26.00	9.83	0.71	0.40	12.37	115.77
	Max	37.91	59.76	16.83	1.78	0.56	22.95	137.65
	Mean	31.75	43.34	13.31	1.23	0.48	18.05	126.69
Treatment 15 ppm As	Min	9.18	13.83	8.00	0.65	0.34	16.08	119.86
	Max	32.78	42.49	15.67	0.96	0.52	27.85	146.54
	Mean	21.69	28.03	11.10	0.80	0.43	22.46	133.77
ANOVA result	G	***	***	***	***	***	***	***
	T	***	***	***	***	***	NA	NA
	G*T	***	***	***	***	***	NA	NA

Significance levels are indicated at P <0.001/*** Treatment and genotype treatment interaction effects were not analysed for As content traits since in control conditions Arsenic is not measured. ND- not determined, NA- not applicable, G- genotype, T- treatment, SPAD- soil plant analysis development, mg kg^{-1}- milligram per kilogram, cm- centimetre, g- gram, As-arsenic, ppm- parts per million and ANOVA- Analysis of variance.

2.3.2. Classification of genotypes for Arsenic tolerance and accumulation

A total of 53 diverse genotypes comprising traditional land-races and released varieties from GSR breeding panel was evaluated for As related traits under 5 ppm, 10 ppm and 15 ppm As concentration for 18 days. Based on the relative phenotypic performance of plants under varying As stress, genotypes were clustered into four

different groups. Genotypes showing high tolerance >80% to As stress were grouped into highly tolerant genotypes and genotypes exhibiting poor tolerance percentage <50% across all the As stress treatment were grouped into highly susceptible genotypes. Genotypes were exhibiting only tolerance in lower concentration As treatment (5 and 10 ppm) as moderately tolerant genotypes and genotypes exhibiting susceptible from 10 ppm As treatment onwards as moderately susceptible genotypes. Further, genotypes were classified into two groups based on As accumulation in the shoots of the genotypes after 18 days of treatment, As shoot includers > 12 mg kg^{-1} and As shoot excluder < 12 mg kg^{-1} (Supplementary data 2). Principal component analysis among the measured As content traits and relative phenotypic traits reveals the pattern of variation among the genotypes, the relationship among individuals and their characteristics. The PC1 explained 24.6%, while, PC2 explained 16.1% variability among genotypes (Figure 2.3). Genotypes, display high susceptibility (M401 and TKM9,) are As shoot includers and genotypes tolerance trend towards restricting As uptake in the shoot (NPT, PSBRc82 and OM997). Genotypes, (Haoannong, BG300 and Phalguna) showed higher susceptibility by excluding As in the shoot. Genotype, Basmati 385 (*Japonica* subspecies) from Pakistan was highly tolerant by including As in the shoot system (Figure 2.3). Genotype (NPT, OM997, Zhong413 and WTR1) was highly tolerant by excluding As in the shoot.

2.3.3. Genotype classification based on grain Arsenic content

Unpolished brown rice of 53 genotypes harvested from the field with soil As content less than 7 mg per kg^{-1} also showed higher variation among the genotypes in grain As accumulation. Among the tested 53 genotypes, As content in grain ranged from 0.12 mg kg-1 in Huang-Hua-Zhan (*Indica* subspecies) from China to 0.48 mg kg-1 in IRAT 109 (*Tropical japonica* subspecies) from Brazil. Based on the concentration on As in the grain, Genotypes were grouped into three groups, low As content rice (<0.2 mg kg-1), moderate As content rice (<0.3 mg kg-1) and High As content rice (>0.3 mg kg-1) rice (Figure 2.4 and Supplementary data 3). Bangladeshi rice genotypes, BR11, and BR24 accumulate >0.4 mg per kg^{-1} of As in the grain. All the *Intermediate* subspecies (Jhona 349, NAN29-2, Basmati 385, BG 300 trend to accumulate moderate As in the grain >0.2 mg per kg^{-1}. Genotypes from *Indica* subspecies (Huang-Hua-Zhan, IR64, GSR IR1-12-Y4-D1-Y2, GSR 12, GSR IR1-4-S5-L1-L1, SACG-4, TME80518, Qi-Si-Ying-Zhan, Te-Huang-Zhan, Zhong413, Basmati and NPT)

accumulate less As (<0.2 mg per kg^{-1}) in the grain. Bangladeshi rice genotypes, BR11 and BR24 accumulate high As in the grain (>0.4 mg per kg^{-1}). High yielding mega variety IR64 accumulate less As < 0.2 mg per kg^{-1} in the grain.

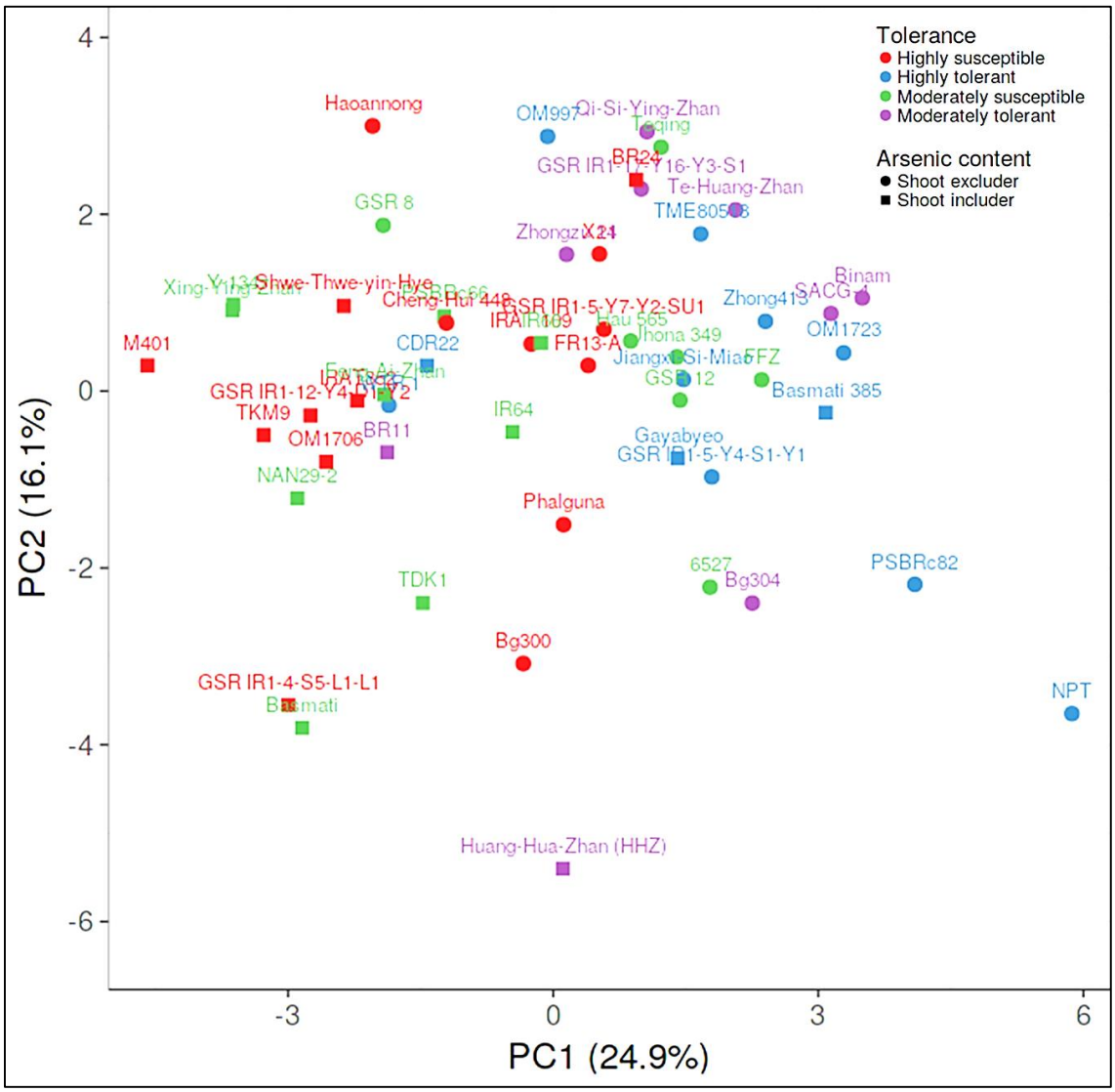


Figure 2.3. Principal component analysis of 53 cultivars of rice (Oryza sativa L.) traits associated with As toxicity stress for 18 days.

Arsenic accumulation in shoot and relative phenotypic variance of morphological traits were considered for principal component analysis. Percentages in parenthesis show those contributed by the first and second principal components.

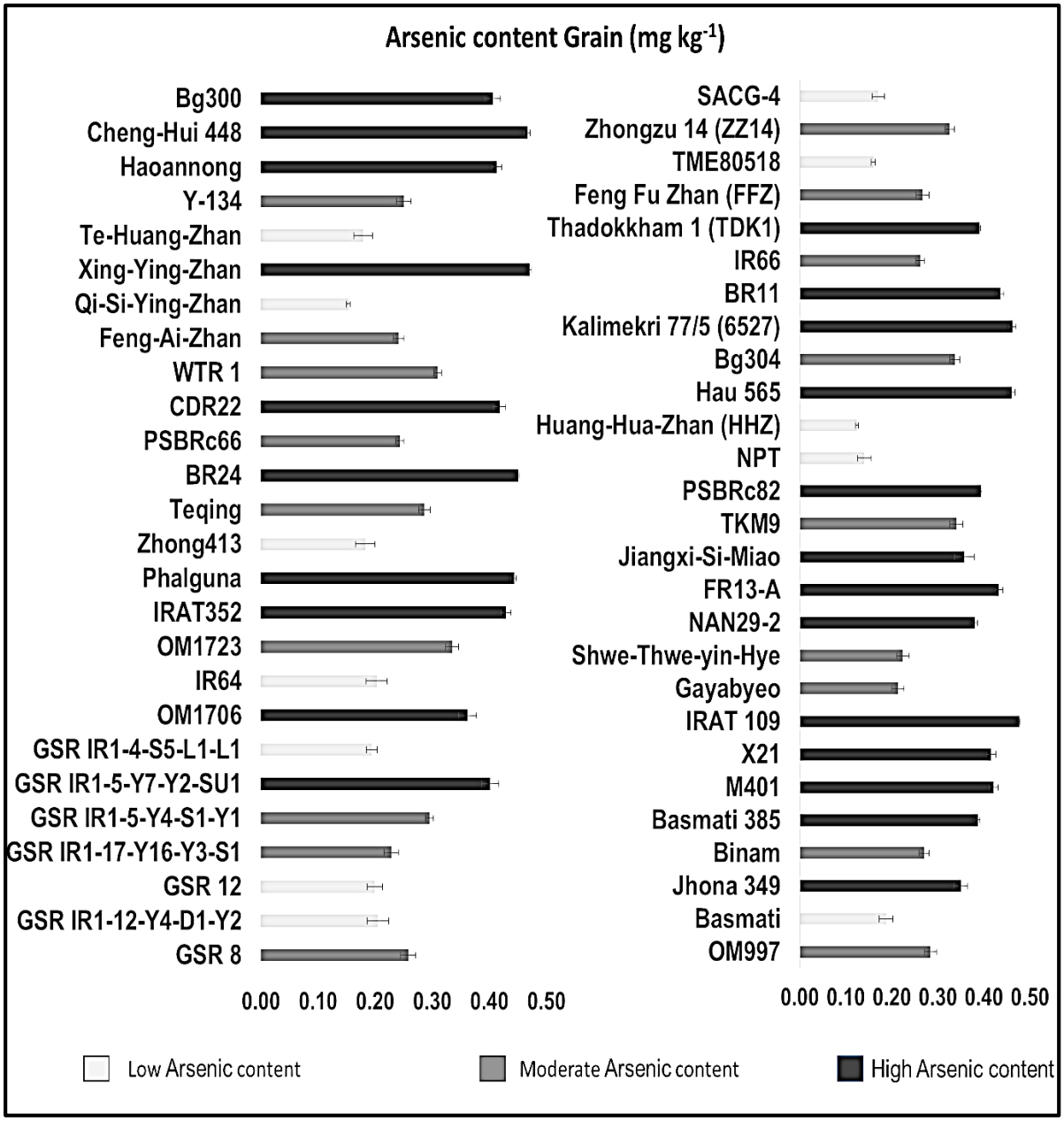


Figure 2.4. Classification of diverse rice genotypes based on As content in unpolished brown rice.

Error bar in the graph represents standard error (n=3) mg Kg^{-1}- milligram per kilogram.

2.4. Developed permanent mapping population

Intended at identifying genomic regions associated with As tolerance and accumulation, two permanent breeding populations were developed through backcross breeding approach (Li and Ali 2017). Among mapping population, one development involved intensive selection strategy by subjecting the BC_1F_2 population derived from an elite cross involving WTR1 as recipient parent and Haoannong as donor parent. High yielding single plants from segregating BC_1F_2 lines were

systematically selected in each advancing generation. First rounds of selection were carried out under rain-fed, low input and irrigated conditions resulting in 576 individual lines. Selected lines form first round underwent three successive selection under varying nutrient conditions (+NPK, 75N,-N, -P, -NP and –NPK) generating 4,848 BC_1F_6 lines. After each generation, the individual selections from the segregating materials were pooled together by keeping their line identity and were screened for all the six conditions. By doing so, the elite selections underwent varied nutrient conditions and stabilised over three rounds of selections and resulted in high yielding stable 230 BC_1F_6 lines (Jewel et al. 2019). Mapping population two were also developed through backcross breeding approach. Eleven donor parents were crossed with single recipient parent WTR1 (Table 5.2). Single plant selections from this population were systematically selected in each generation from segregating BC_1F_2 population (Ali et al. 2018b). First rounds of selection under rain-fed, low input and irrigated conditions give rise to 730 lines. In second round compound selection were carried out for tolerance under (salinity, submergence, anaerobic germination and tungro) and plant yield under (irrigated, low input and rain-fed condition) give rise to 1151 lines. In the third round of selection, selected 1151 individuals were screened for yield under irrigated, low input and rain-fed condition produce 2242 lines. Out of 2242 lines, only 564 high yielding lines and 12 parents were considered for genotyping.

2.5. Mapping genomic regions associated with Arsenic stress in backcross breeding population

2.5.1. Performance of parental lines

Exposure to 10 ppm of As for 18 days induced a significantly negative response in parents of the mapping population. WTR1 showed more tolerance than Haoannong in terms of chlorophyll content and biomass accumulation. With regard to root length and plant height, Haoannong showed higher tolerance. The parents WTR1 and Haoannong showed highly significant differences in accumulating As between shoot and root tissue (Figure 2.5).

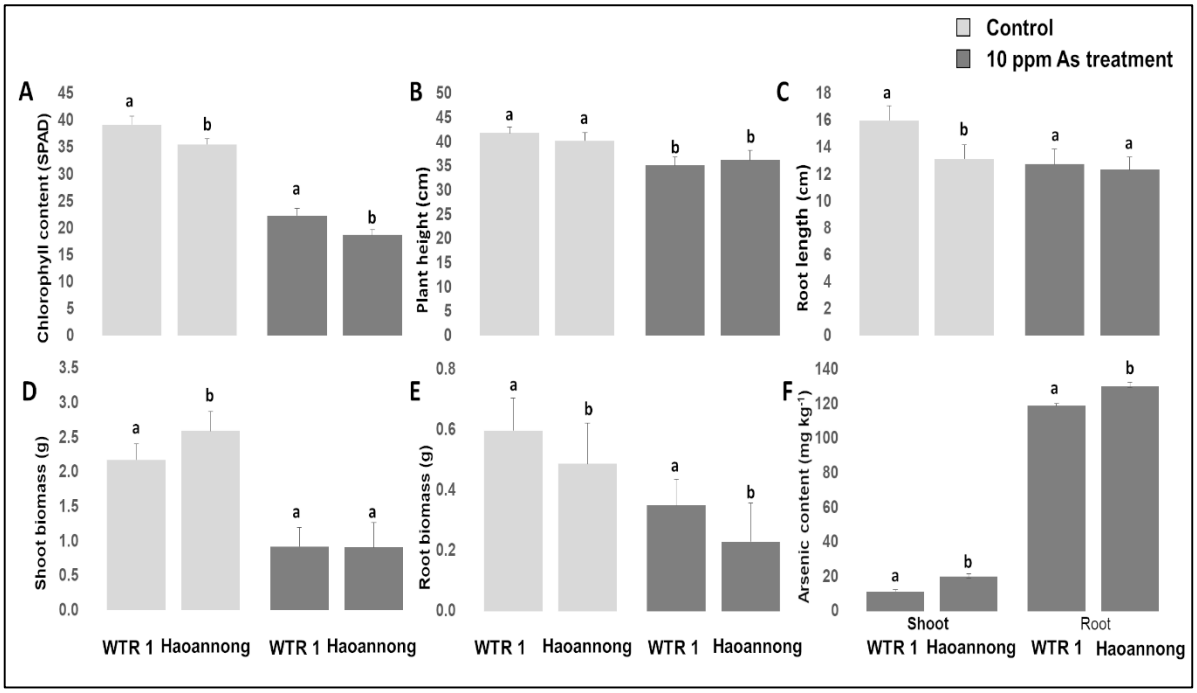


Figure 2.5. Phenotypic trait variation in parents of mapping population under As toxicity stress for 18 days.

(A) Chlorophyll content; (B) plant height; (C) root length; (D) shoot dry weight; (E) root dry weight; (F) Arsenic content in shoot and root tissue. Vertical bars represent standard errors of means (n = 3). Different letters above the data points indicate significant differences between genotypes by LSD test (P <0.05).

2.5.2. Performance of mapping population under Arsenic stress

When averaged over all lines, the mean values of shoot dry weight, plant height, and chlorophyll content decreased considerably in the As stress, except for root biomass and root length (Table 2.2). As content in BRILs ranged from 9.10 mg kg^{-1} (AsG-195) to 20.99 mg kg^{-1} (AsG-227) and 68.62 mg kg^{-1} (AsG-185) to 197.10 mg kg^{-1} (AsG-226) in shoot and root, respectively. Compared with the parents, 15 lines showed the lowest uptake of As concentration (≤10.00 mg kg^{-1} in shoots and ≤80.00 mg kg^{-1} in roots). In particular, two genotypes, AsG-187 (*GSR IR2-1-R14-N2-N5-N21-N42*) and AsG-189 (*GSR IR2-1-Y4-N4-N2-N3-N2*), showed better performances in low uptake of As content in roots and shoots (Supplementary data 4). Most of the traits measured in the control and treatment conditions appeared to be normally distributed (Figure 2.6). The overall results indicated that adequate genetic variability is present in the populations studied for mapping QTLs.

Table 2.2: Descriptive statistics and ANOVA results for different phenotypes.

	Control			Treatment 10 ppm As			ANOVA result		
Trait	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**G**	**T**	**G*T**
Chlorophyll content (SPAD)	29.93	40.72	35.31	16.15	31.45	23.27	***	***	NS
Plant height (cm)	39.33	63.17	48.25	23.13	43.50	33.31	NS	***	***
Root length (cm)	10.83	19.83	15.86	9.00	18.17	13.00	***	***	***
Shoot dry weight (g)	0.91	2.22	1.56	0.82	1.15	0.98	***	***	***
Root dry weight (g)	0.31	0.72	0.53	0.20	0.70	0.45	***	***	NS
Arsenic content of shoots (mg kg^{-1})	ND	ND	ND	9.10	20.99	15.48	***	NA	NA
Arsenic content of roots (mg kg^{-1})	ND	ND	ND	68.62	197.10	130.36	***	NA	NA

Significance levels are indicated at P <0.001/*** Treatment and genotype treatment interaction effects were not analysed for As content traits since in control conditions As is not measured. ND: not determined, NS: non-significant, NA: not applicable, G: genotype, T: treatment.

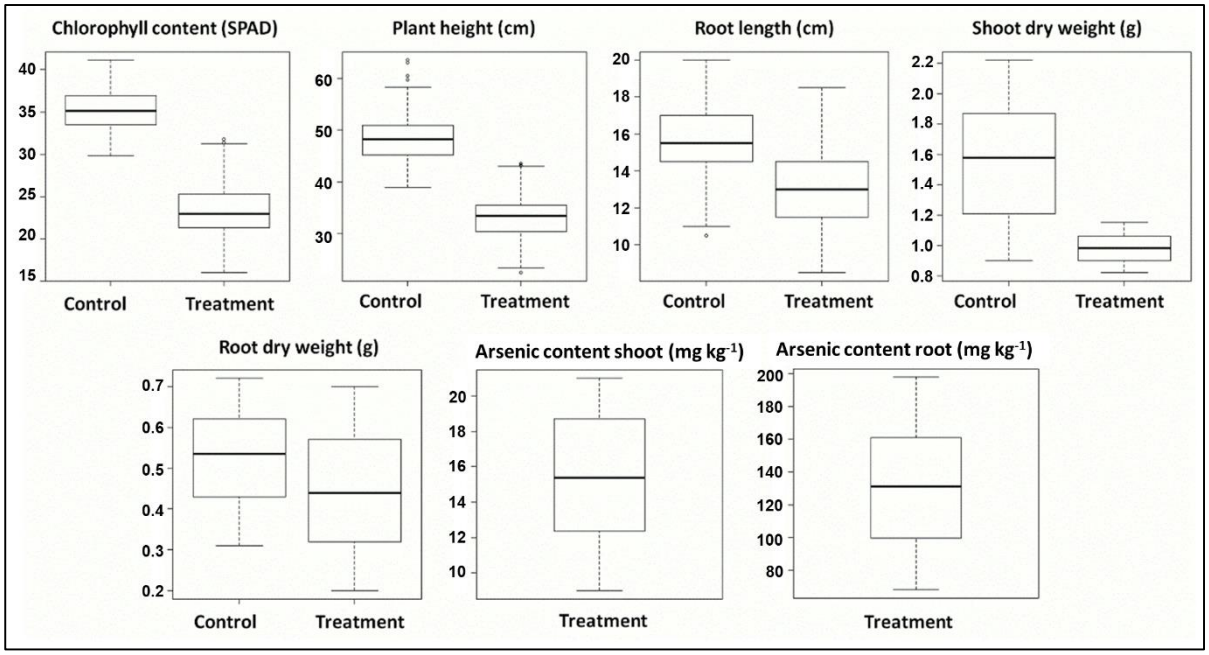


Figure 2.6. Box plot showing the distribution of measured As-related stress in the single backcross breeding population.

Control with no arsenic and treatment with 10 ppm arsenic substitution for 18 days. The thick line in the middle of the box is the median of the distribution while the lower and upper boundaries represent first and third quartile, respectively. Lower and upper whiskers are calculated based on 1.5 times the inter-quartile range.

2.5.3. Correlation among measured traits

The study of correlation coefficient among various measured As-related traits is of great importance to predict the relationship between complex quantitative As content traits and simple quantitative morphological traits associated with As toxicity. Arsenic content in shoots, a significantly negative correlation with relative plant height, chlorophyll content, and root length (Figure 2.7). Arsenic content in roots, there was a significant negative correlation with relative root length and plant height. A significant and positive correlation coefficient was recorded between As content in shoots and roots, relative root length and relative shoot length, and relative shoot biomass and relative root biomass.

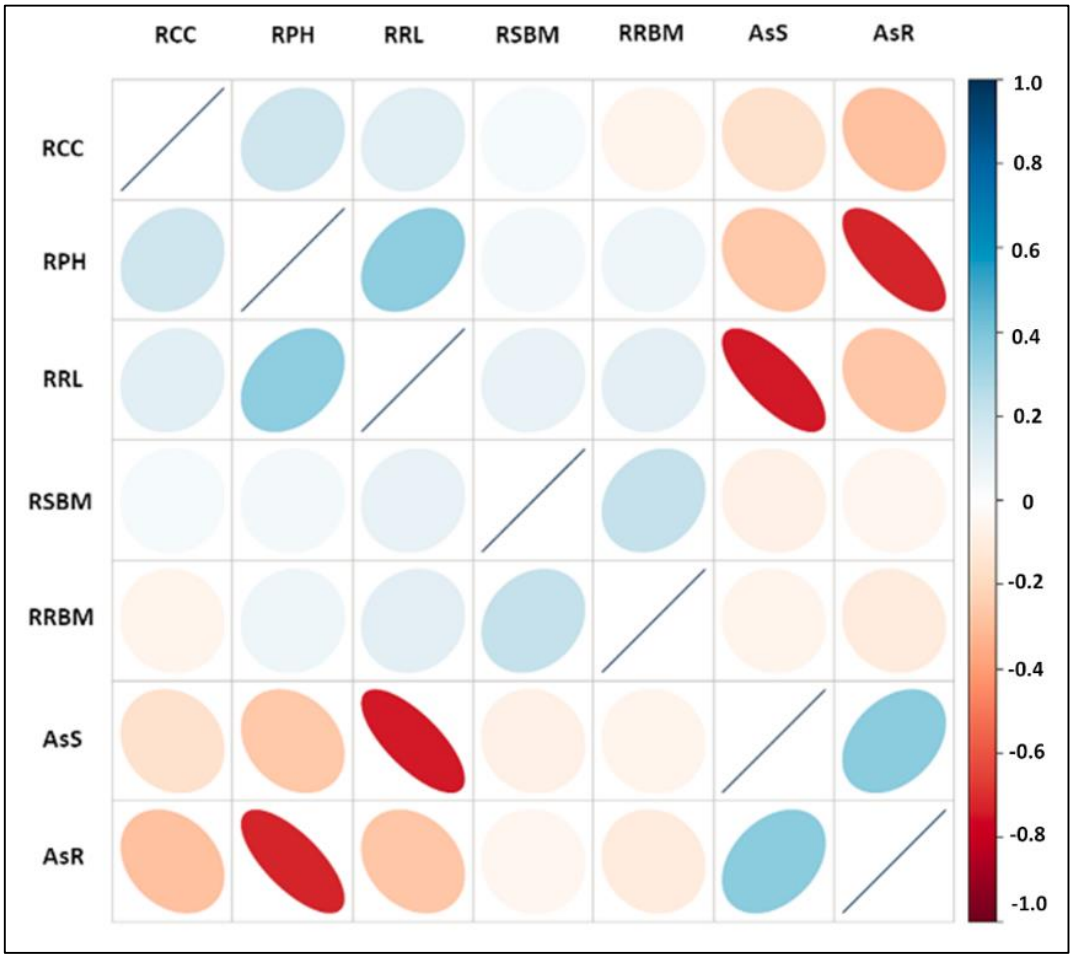


Figure 2.7. Heat map is showing the correlation coefficient of As-related traits. Positive correlations are displayed in blue and negative correlations in red colour. Colour intensity and the size of the circle are proportional to the correlation coefficients (P< 0.01). RCC: relative chlorophyll content; RPH: relative plant height; RRL: relative root length; RSBM: relative shoot biomass; RRBM: relative root biomass; AsS: As content in shoots; AsR: As content in roots.

2.5.4. Single nucleotide polymorphism markers generated by 6K SNP-array

A customised 6K SNP-Beadchip was used. It consists of 4606 SNP markers, which detected 1068 polymorphic sites between the parents and BC_1F_6 BRILs. A total of 704 polymorphic SNPs remained after pairwise comparison of SNPs between the parents for missing and heterozygous SNPs. These 704 SNPs were unevenly distributed across the genome, ranging from 84 SNPs on chromosome 4 to 41 on chromosome 9, with an average spacing of ~524 kb between the SNPs, ranging from 393.9 kb on chromosome 7 to 797.4 kb on chromosome 2 (Figure 2.8). There were 42 significant gaps (>2 Mb) across the genome in the distribution of SNPs generated by the 6K SNP-array as a result of monomorphic SNP markers shared between parents in these regions. Gaps ranging from 4 Mb on chromosome 2 to 8 Mb on chromosome 11 were observed in the generated physical map of the rice genome using the 704 SNP marker data (Figure 2.9). These 704 filtered SNPs were used to investigate the association between markers and As-related traits.

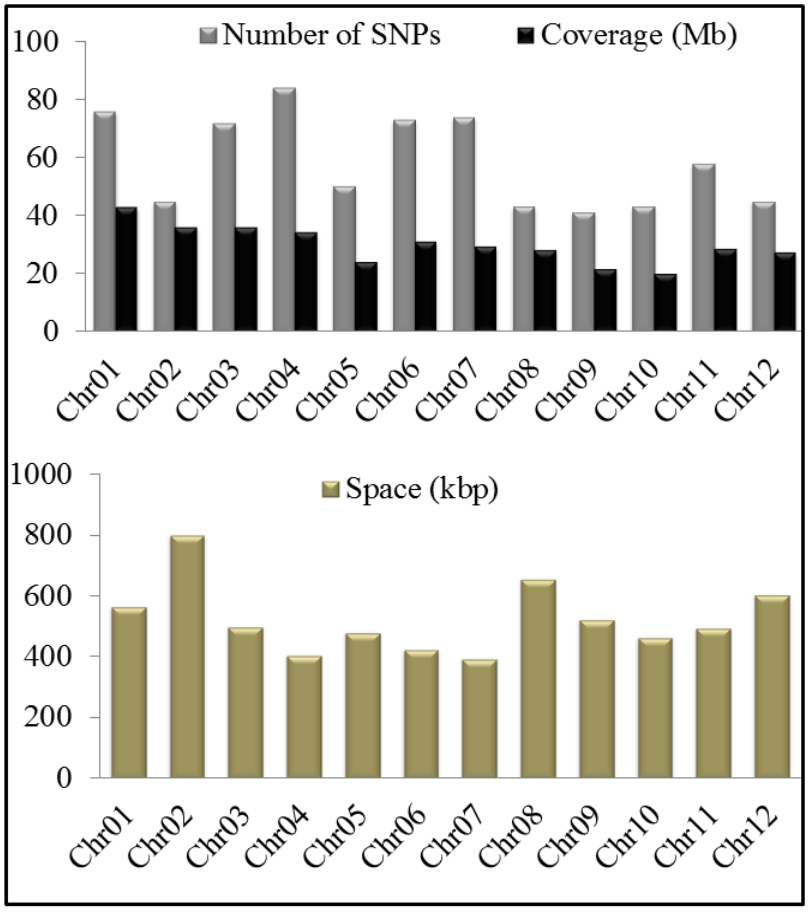


Figure 2.8. Distribution of SNPs in the chromosome.

a) The number of polymorphic SNPs and their coverage in chromosome b) Spacing between the SNP markers. Chr: Chromosome; Mb: Megabase pair; kbp: Kilobase pair

2.5.5. QTLs associated with Arsenic-related traits

In total, 40 SNPs showed significant marker-trait association for the As-related traits (Table 2). Nine QTLs were defined by assuming that closely linked significant markers are in the QTL region. Among the nine QTLs, six were identified for As content in shoots, two were identified for As content in roots, and one tolerant QTL was identified for relative chlorophyll content. QTLs for As content in shoots (*qAsS2, qAsS5.1, qAsS5.2, qAsS6, qAsS9.1,* and *qAsS9.2*) were mapped on chromosomes 2, 5, 6, and 9, respectively. QTLs for As content in roots (*qAsR8.1* and *qAsR8.2)* were mapped on chromosome 8 and relative chlorophyll content QTL *qRChlo1* was mapped on chromosome 1 (Figure 2.9). All the QTLs detected (P <0.01) had a –log *p*(F) ≥3.3 and phenotypic variance explained by the QTLs ranged from 8.6% to 12.6%, with QTLs *qAsS5.2, qAsS6,* and *qAsR8.1* explaining major phenotypic variance of more than 10% and QTLs *qRChlo1, qAsR8.2, qAsS2, qAsS5.1, qAsS9.1,* and *qAsS9.2* explaining minor phenotypic variance of less than 10% (Table 2.3).

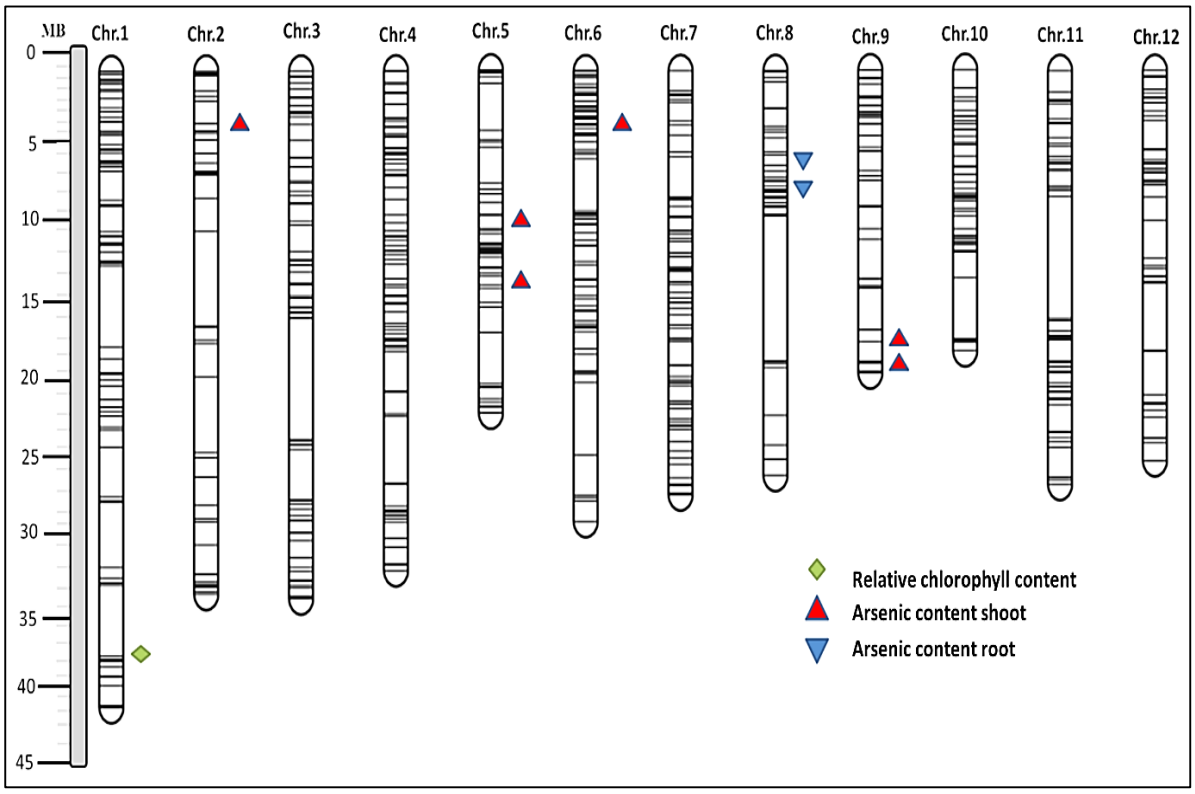


Figure 2.9. Chromosomal distribution of polymorphic single nucleotide polymorphisms (SNPs) and detected quantitative trait loci (QTLs).

QTLs were located on the chromosome based on the physical position of the SNP marker.

Table 2.3: QTLs associated with As-related traits after 18 days of 10 ppm As stress in WTR1 × Haoannong population by single marker regression

QTL[a]	Trait	Chr	Marker[b] position	Associated[c] marker	d-log p(F)[d]	R2 (%)[e]	Additive[f] effect	Tolerance[g] allele
qRChlo1	Relative	1	39.2829	66SNP_1_39282883	3.78	8.62	4.07	WTR1
	Chlorophyll	1	39.3692	67SNP_1_39369209	3.78	8.62	4.07	WTR1
	content	1	39.4208	68SNP_1_39420824	3.78	8.62	4.07	WTR1
qAsR8.1	Arsenic	8	6.0577	487SNP_8_6057678	4.65	10.51	-11.79	WTR1
qAsR8.2	content root	8	7.854	492SNP_8_7854002	3.82	8.72	-10.67	WTR1
qAsS 2	Arsenic	2	4.3429	88SNP_2_4342883	4.33	9.81	-1.13	WTR1
	content	2	4.9307	89SNP_2_4930742	4.14	9.41	-1.10	WTR1
	shoot	2	5.8303	90SNP_2_5830265	4.03	9.18	-1.09	WTR1
		2	6.4799	91SNP_2_6479920	4.17	9.48	-1.11	WTR1
		2	7.0233	92SNP_2_7023295	4.17	9.48	-1.11	WTR1
		2	7.0767	93SNP_2_7076671	4.33	9.82	-1.12	WTR1
		2	7.1037	94SNP_2_7103684	4.33	9.82	-1.12	WTR1
		2	7.1708	95SNP_2_7170842	4.33	9.82	-1.12	WTR1
		2	7.2388	96SNP_2_7238793	4.33	9.82	-1.12	WTR1
		2	7.2775	97SNP_2_7277487	4.17	9.48	-1.11	WTR1
qAsS 5.1	Arsenic	5	10.8863	287SNP_5_10886331	3.72	8.50	-1.05	WTR1
	content	5	13.3255	288SNP_5_13325546	3.72	8.50	-1.05	WTR1
	shoot	5	13.7687	289SNP_5_13768744	3.72	8.50	-1.05	WTR1
		5	14.0797	290SNP_5_14079677	3.72	8.50	-1.05	WTR1
		5	14.644	291SNP_5_14643984	4.12	9.37	-1.16	WTR1
qAsS5.2		5	15.4693	292SNP_5_15469279	4.03	9.17	-1.09	WTR1
		5	15.556	293SNP_5_15556017	4.03	9.17	-1.09	WTR1
		5	16.4598	294SNP_5_16459802	4.03	9.17	-1.09	WTR1
		5	16.5851	295SNP_5_16585060	4.03	9.17	-1.09	WTR1
		5	16.8086	296SNP_5_16808642	5.66	12.64	-1.36	WTR1
qAsS6	Arsenic	6	0.4008	329SNP_6_400753	4.11	9.34	1.38	HAN
	content	6	0.6469	330SNP_6_646915	4.40	9.98	1.47	HAN
	shoot	6	0.8342	332SNP_6_834170	4.11	9.34	1.38	HAN
		6	1.5219	335SNP_6_1521855	3.58	8.19	1.38	HAN
		6	1.768	336SNP_6_1768006	5.32	11.93	1.49	HAN
		6	1.9284	337SNP_6_1928403	4.01	9.14	1.28	HAN
		6	1.9783	338SNP_6_1978288	4.01	9.14	1.28	HAN
		6	2.0107	339SNP_6_2010737	4.51	10.21	1.33	HAN
		6	2.0256	340SNP_6_2025629	4.51	10.21	1.33	HAN
qAsS9.1	Arsenic	9	18.3666	551SNP_9_18366555	3.88	8.85	-1.09	WTR1
	content	9	18.3891	552SNP_9_18389054	3.88	8.85	-1.09	WTR1
	shoot	9	19.2081	553SNP_9_19208050	3.77	8.59	-1.06	WTR1
qAsS9.2		9	20.587	555SNP_9_20587039	3.98	9.06	-1.08	WTR1
		9	21.2154	557SNP_9_21215424	3.67	8.38	-1.04	WTR1
		9	21.3489	558SNP_9_21348882	4.37	9.90	-1.13	WTR1

[a] Closely linked markers are assumed as the same QTL, [b] Physical position of markers on chromosomes, [c] Marker associated with QTL, [d] F-statistical analysis indicates an association between markers and trait, [e] Proportion of phenotypic variance explained, [f] Positive/negative values indicate that WTR1/Haoannong can increase trait values and [g] Tolerance allele provided by parental line.

2.5.6. Selection of candidate genes associated with Arsenic-related traits

In the confidence interval of the identified QTLs, 35767 and 440 SNPs were identified between the parents in the Rice SNP-Seek Database and tGBS® dataset, respectively. Among those identified SNPs, 35% of the SNPs showed polymorphism between the parents in 1309 genes, and most of these polymorphisms (93.9%) were synonymous mutations, with non-synonymous mutations (6.1%) between the parents in 676 genes (Table 2.4 and Supplementary data 4). Among those 676 genes, 35 genes were associated with metal transport, response, and metal homeostasis and thus considered as the most likely candidate genes (Supplementary data 5). Expression data of these genes were retrieved from a study by Yu et al. (2012) using the web application Rice Expression Database (Xia et al. 2017). Twenty-five genes out of the 35 most likely candidate genes showed expression responses to the sodium arsenite (As^{III}) stress treatment according to expression data (Yu et al. 2012). Among the 25 genes showing expression responses, three genes were up-regulated, and 22 showed significant down-regulation (Figure 2.10).

Table 2.4: SNPs identified in the QTL regions using tGBS and 3K RGP.

QTLs	Chr	No. of SNPs in tGBS	No. of SNPs in 3K RGP	Polymorphic SNPs	Non-synonymous SNPs	Total no. of genes	Genes with non-synonymous SNPs
qRChlo1	1	4	479	202	63	20	20
qAsR8.1	8	34	2761	757	148	47	27
qAsR8.2	8	65	3070	1110	179	61	42
qAsS2	2	79	11705	4542	821	382	225
qAsS5.1	5	46	2346	677	108	94	45
qAsS5.2	5	103	2166	976	192	113	67
qAsS6	6	86	6789	1932	357	240	98
qAsS9.1	9	11	6154	2214	309	326	141
qAsS9.2	9	12	297	142	16	26	11
Total		440	35767	12552	2193	1309	676

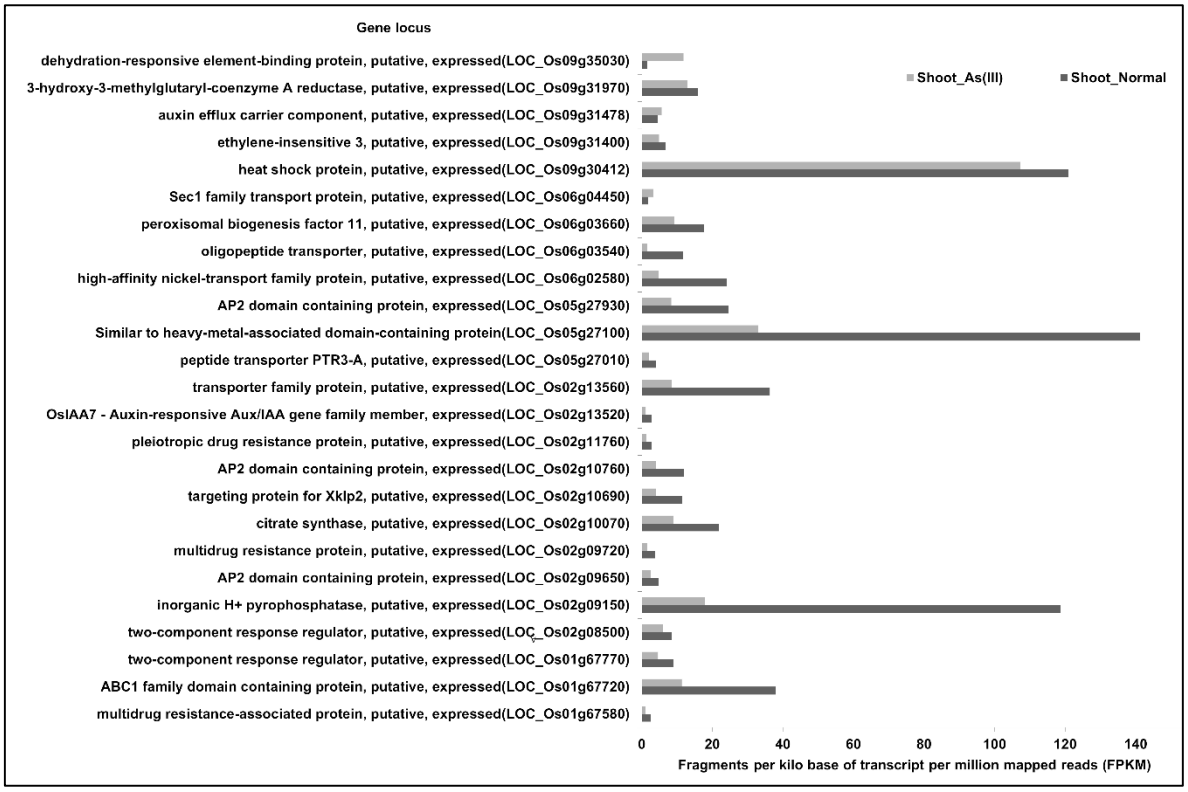


Figure 2.10. Transcriptional regulation of genes contained in identified QTL regions. Only genes containing non-synonymous SNPs showing significant expression response under $As^{(III)}$ stress are plotted. Gene expression data adopted from Yu et al. (2012).

2.6. Exploiting tGBS based SNP genotyping in the interconnected breeding population

2.6.1. tGBS® based SNP genotyping on population.

The tGBS® analysis was used for SNP-typing a rice diversity panel comprising of 11 donor parents, one common recurrent parent and 564 BC_1F_5 lines (Supplementary data 7). The 943.4 M raw tGBS® sequencing reads used in the current study were generated using 10 Ion Proton runs. After trimming low-quality bases, 881.6 M reads, and 87.8% of base pairs were retained. Approximately 80.9% and 65.7% of the trimmed reads could be aligned non-uniquely and uniquely, respectively. Using the reads from the 576 samples that uniquely aligned to the reference genome, 794,297 polymorphic sites were identified after interrogating 2,679,180 bases that have ≥five reads in at least 50% of the samples. After filtering, a low-missing data set (LMD50) was identified (Table 2.5).

Table 2.5: SNP summary and parent genotype categories ratios.

Sub- Pop*	Donor Parent (DP) and Source Country	No. of lines	LMD50 SNPs
1	Haoannong (DP1)-China	120	4,669
2	Cheng-Hui 448 (DP4)-China	67	5,968
3	Feng-Ai-Zan (DP5)-China	42	2,284
4	Y 134 (DP6)- China	34	4,035
5	Zhong 413 (DP7)-China	56	3,149
6	Khazar (DP8)-Iran	29	3,962
7	BG 300 (DP9)-Srilanka	60	5,226
8	OM 997 (DP10)-Vietnam	55	5,921
9	Basmati-385 (DP12)-Pakistan	31	5,995
10	M 401 (DP17)-USA	33	7,045
11	X 21 (DP19)-Vietnam	37	6,985
Total	11 DPs	564	55,239

*All 11 sub-populations have a common recipient parent, Weed Tolerant Rice-1

2.6.2. Polymorphic SNP analysis

Of the total LMD50 SNPs discovered in all 11 subpopulations (Table 2.5), ~40% were located in genic regions (Figure 2.11). Exonic or CDS regions contained 4,784 non-synonymous SNPs and 3,373 synonymous SNPs. As expected, the majority of SNPs 21% located in the genic region were intronic. The LMD50 SNPs were not uniformly distributed across chromosomes (Figure 2.12). SNP distribution varied within chromosomes for all the 11 subpopulations. The analysis revealed significant SNP hotspots across all chromosomes and genomic regions where no SNPs were identified. Sub-population 1 consisted of 122 samples (120 lines and the two parents, Table 2.5). Using aligned reads from the 122 lines 4,669 high-quality SNPs (LMD50) were identified (Table 2.5).

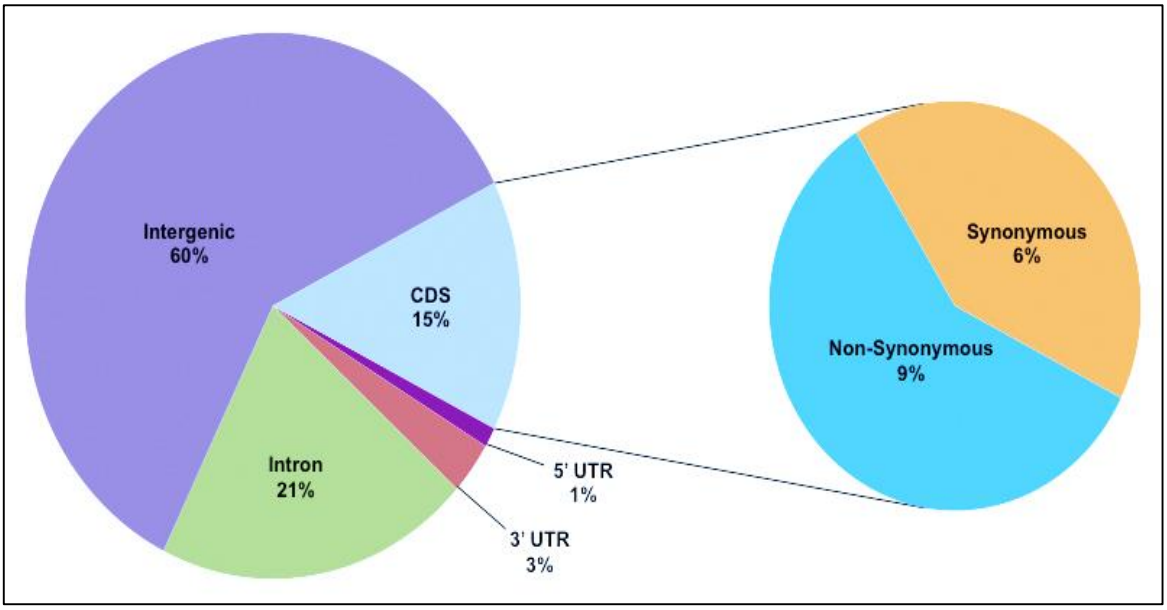


Figure 2.11. Annotation of LMD50 SNPs.

Intergenic and genic proportions of identified SNPs. CDS- coding region of the genome.

2.6.3. Distribution of LMD50 SNP in the genome

The number of LMD50 SNPs per sample that are homozygous for the recurrent parent allele, homozygous for the donor parent allele, and heterozygous and missing are shown in (Figure 2.13), which also illustrates the proportion of the SNPs per sample that are homozygous for the recurrent parent allele, homozygous for the donor allele, or heterozygous among non-missing genotypes. Key figures for the LMD50 SNPs for the other ten sub-populations are provided in the supplemental information (Supplementary figure 2.1 to 2.3). The LMD50 SNPs for each sample in sub-population one were plotted by their physical ordering on concatenated chromosomes, providing a clear visualisation of genotypic patterns (Figure 2.15). This display of LMD50 SNPs allows clustering of SNPs that are heterozygous or homozygous for the donor parent allele in specific chromosomal regions. Similar displays of LMD50 SNPs for the other ten sub-populations are provided in Supplementary figures (Supplementary figure 2.4 to 2.6).

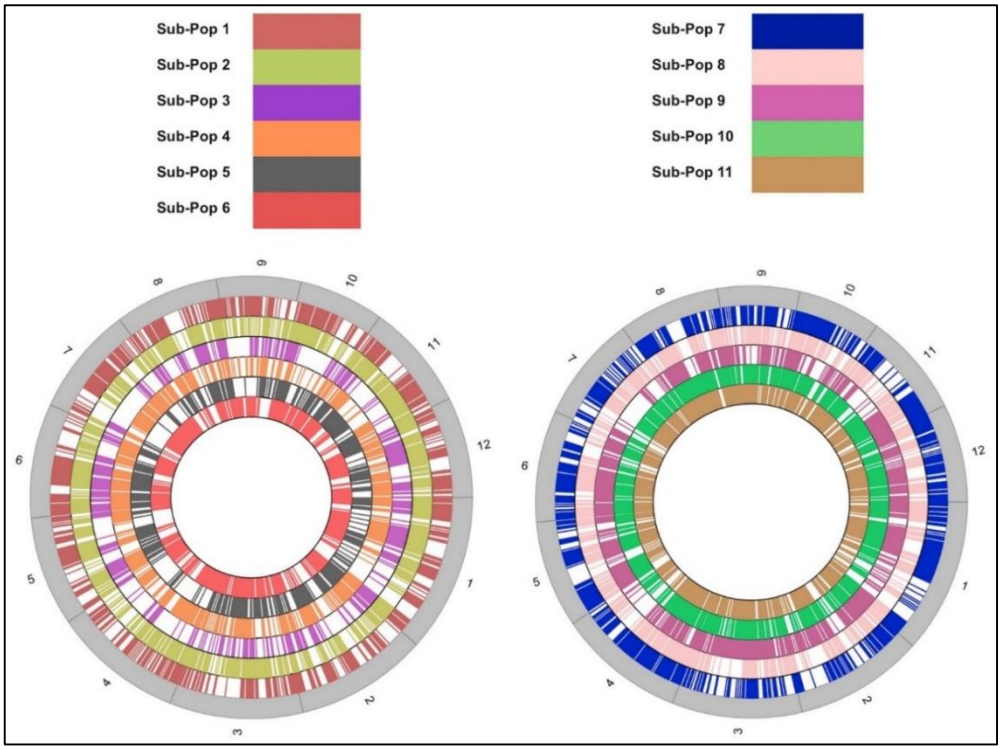


Figure 2.12. Genomic distribution of LMD50 SNPs.

The size of each chromosome is based on the Rice Genome Annotation Project (MSU7; http://rice.plantbiology.msu.edu/index.shtml) database.

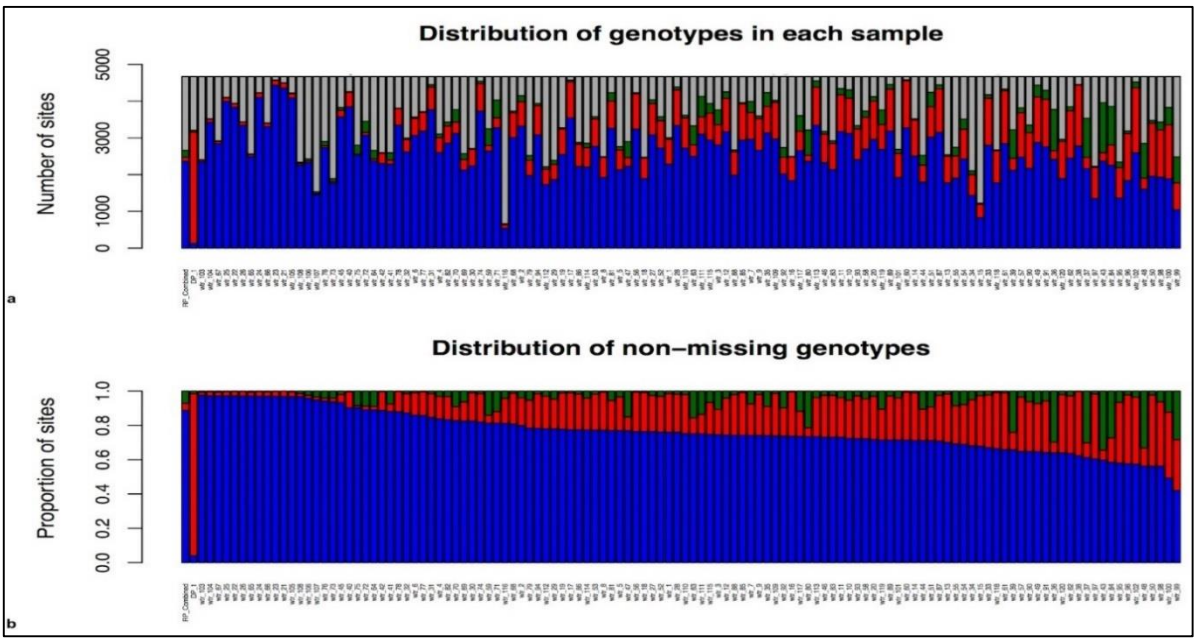


Figure 2.13. LMD50 SNPs genotype summary by sample for subpopulation 1.

a) Including missing SNP calls; b) considering only non-missing SNP calls. (Red colour represents homozygous donor parent alleles, blue represents homozygous recurrent parent alleles and green for heterozygous).

2.6.4. Phylogenetic analysis of developed breeding population

The phylogenetic analysis based on the LMD50 SNPs revealed a clear differentiation of the 576 genotypes into 12 distinct groups in a neighbour-joining tree (Figure 2.14). The 12 groups correspond to the 11 donor parents of each of the 11 subpopulations (Table 2.5) and the recurrent parent.

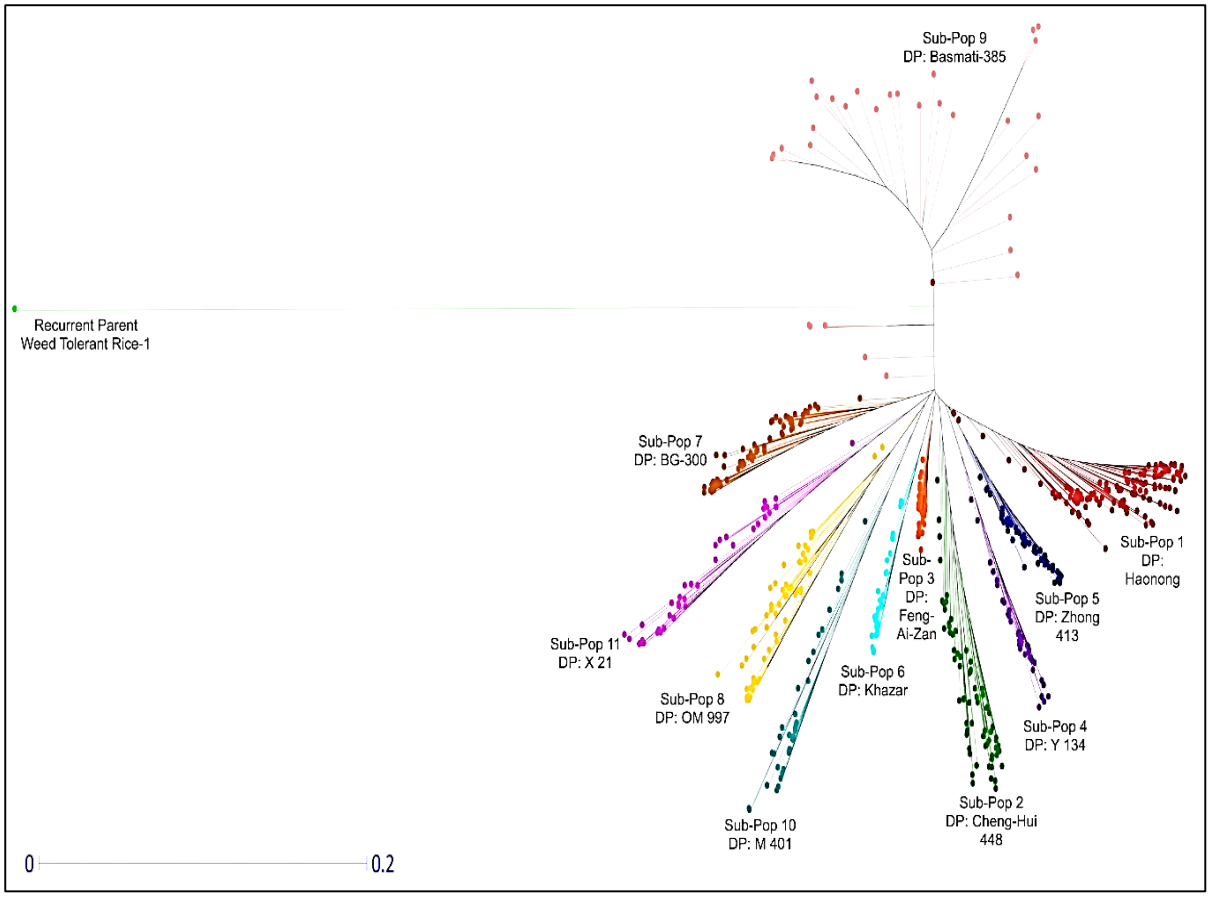


Figure 2.14. Diversity and phylogenetic patterns among 11 subpopulations. Neighbour-joining tree based on LMD50.

2.6.5. The widespread occurrence of selective sweeps across the population

It is notable that on chromosome 4 of sub-population 1 (Figure 2.14), considerable evidence of selection sweep exists, implying the increased frequency of donor alleles in response to severe selection pressure under abiotic stress. Similar occurrences of selective sweeps are evident on various chromosomes in other introgression populations and are indexed in supplementary figures (Supplementary figure 2.4 to 2.6)

Figure 2.15. Chromosomal representation of SNPs (LMD50) sub-population 1. Red colour indicates donor parent allele, the blue colour represents recurrent parent allele, and green colour represents heterozygous loci.

2.6.6. Allele frequency in the population

The recurrent parent allele frequency was also plotted by SNPs and selective window scanning (window size: 10 SNPs and step size: 5 SNPs was used) (Figure 2.15). Additionally, to visualise the un-expected high donor parent allele frequency presumably implying to selective sweep on chromosome 4 in comparison with chromosome 10, same window scanning parameters were used with average donor parent allele frequency of each window was plotted (Figure 2.16). Recurrent parent

allele frequency for each population by SNPs and selective window scanning (window size 10 SNPs and a step size of 5 SNPs was used) were plotted and aggregated for a broader visualisation. The average recurrent parent allele frequency ranged between 0.7-0.8 for those populations (Figure 2.17 and 2.18).

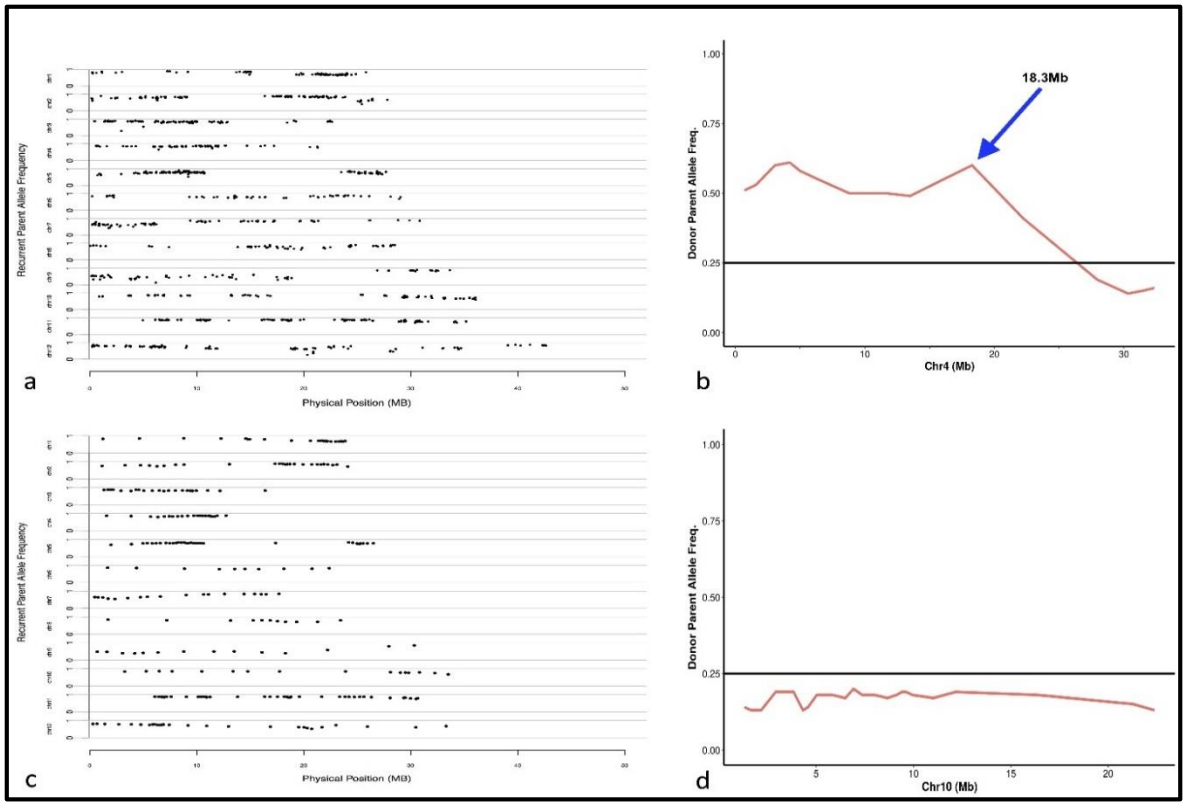


Figure 2.16. Allele frequencies in subpopulation 1.

a) Recurrent parent allele frequency plotted by SNPs (LMD50) among 12 chromosomes. b) Recurrent parent allele frequency plotted by window scanning method (window size of 10 SNPs and a step size of 5 SNPs was used). c) Donor parent allele frequency on chromosome 4 (window size of 10 SNPs and a step size of 5 SNPs was used, the black line (0.25) represents average donor parent allele frequency among all population's parental LMD50 SNPs). d) Donor parent allele frequency on chromosome 10.

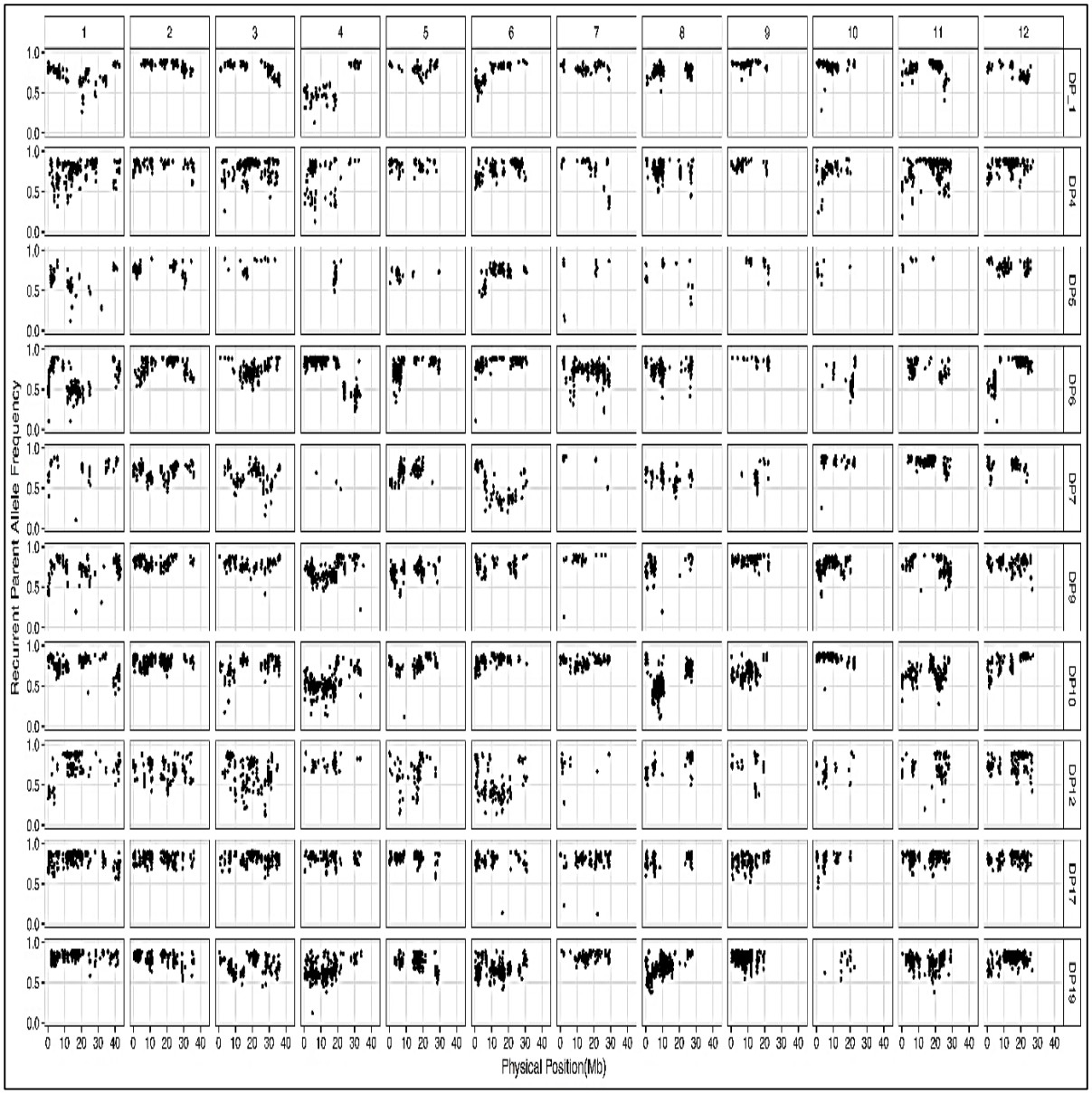


Figure 2.17. Recurrent parent allele frequencies of introgression populations by SNPs (LMD50).

Frequency distribution of 12 chromosomes represented for each population titled with its donor parent (DP) identity.

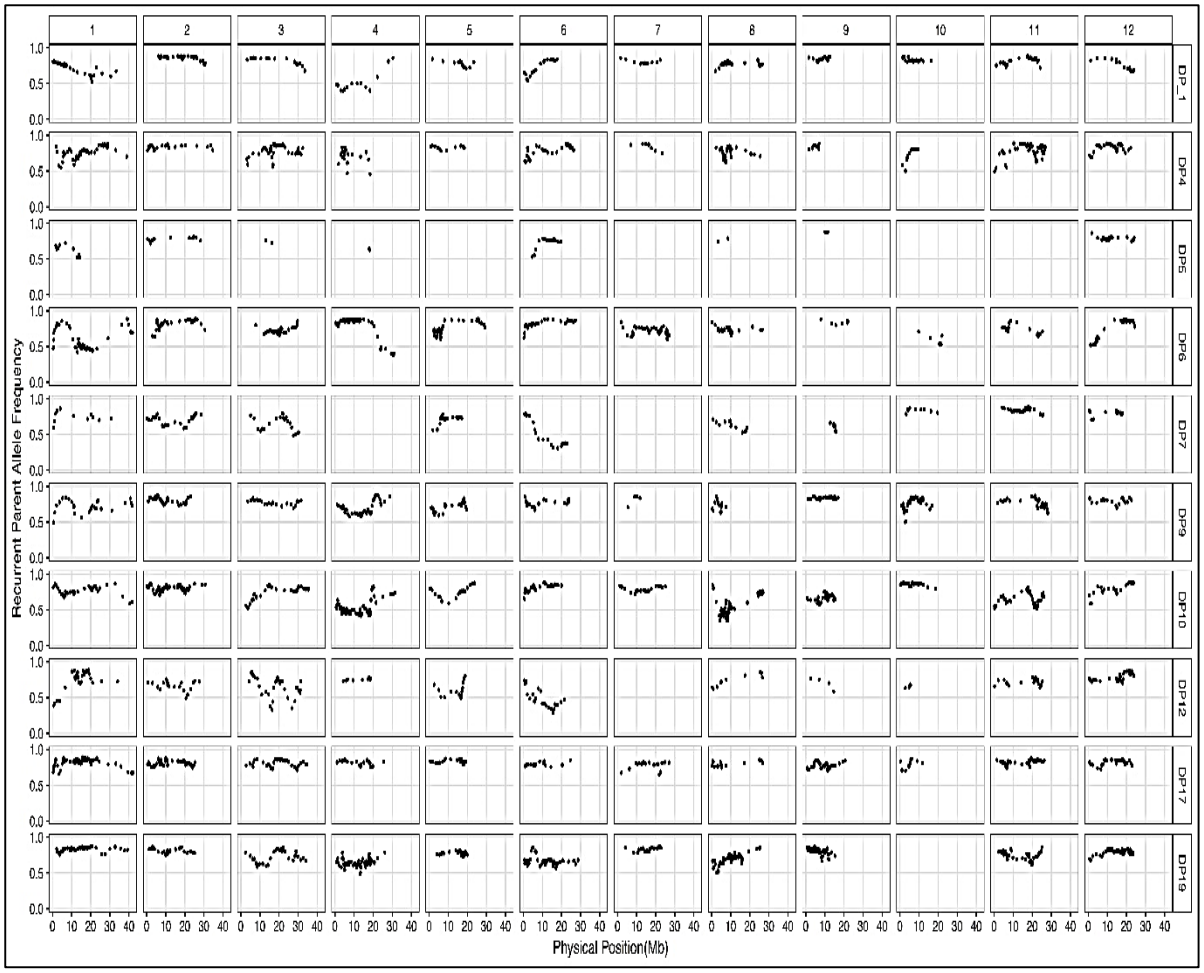


Figure 2.18. Recurrent parent allele frequencies of introgression populations by window scanning method.

Frequency distribution of 12 chromosomes represented for each population titled with its donor parent (DP) identity. The window size of 10 SNPs and a step size of 5 SNPs were used.

2.7. Genomic regions associated with As toxicity stress in interconnected backcross breeding population

2.7.1. Mapping population for As stress experiment

Breeding population panel consists of 11-backcross breeding populations (BC_1F_5), comprising a set of 564 diverse lines and 12 parents. All the 11 breeding population share common recurrent parent WTR1. Based on the contrasting tolerance pattern from the breeding panel screening experiment, four populations from the 11-backcross breeding populations were selected to identify QTLs associated with As-related traits (Table 2.6). The combined breeding population includes 34 lines from cross WTR1 × Y 134, 56 lines from cross WTR1 × Zhong 413, 60 lines from cross WTR1 × BG300

and 55 lines from cross WTR1 × OM997. In total, the mapping population consisting of 210 individual including five-founder parents were used for further analysis (Supplementary data 7).

Table 2.6: Classification of parents for As stress in combined BRIL population.

Parents*	Germination assay	Hydroponic screening	Shoot As content	Grain As content
Haoannong (DP1)	Tolerant	Susceptible	Excluder	High
Cheng-Hui 448 (DP4)	Susceptible	Susceptible	Excluder	High
Feng-Ai-Zan (DP5)	Tolerant	Susceptible	Includer	High
Y 134 (DP6)	Tolerant	Susceptible	Includer	Moderate
Zhong 413 (DP7)	Tolerant	Tolerant	Excluder	Low
Khazar (DP8)	Tolerant	Tolerant	Includer	Moderate
BG 300 (DP9)	Tolerant	Susceptible	Excluder	High
OM 997 (DP10)	Tolerant	Tolerant	Excluder	High
Basmati-385 (DP12)	Susceptible	Tolerant	Includer	High
M 401 (DP17)	Tolerant	Susceptible	Includer	High
X 21 (DP19)	Susceptible	Susceptible	Excluder	High
WTR 1(Rp)	Tolerant	Tolerant	Excluder	Moderate

* Classification was based on breeding panel As screening experiment.

2.7.2. Performance of mapping population under Arsenic stress

Hydroponic phenotyping experiment was set up to test the influence of varying As stress treatment on mapping population. Presence of As in the nutrient solution for 18 days induced visible toxicity symptoms in the population, plant height and greenness was reduced considerably with the increasing As concentration in the treatment (Figure 2.18). When averaged over all lines, the mean value of biomass, plant height, and chlorophyll content decreased considerably with the increasing concentration of As in the treatment (Figure 2.19). Mapping population has an average relative tolerance of 82% in 5 ppm As, 65.8% in 10 ppm As and lowest 49.6% in 15 ppm As treatment (Supplementary data 8). As content in shoot and root trend to increase with

the increasing As concentration in the treatment. All the measured phenotypes were summarised in the table (Table 2.7).

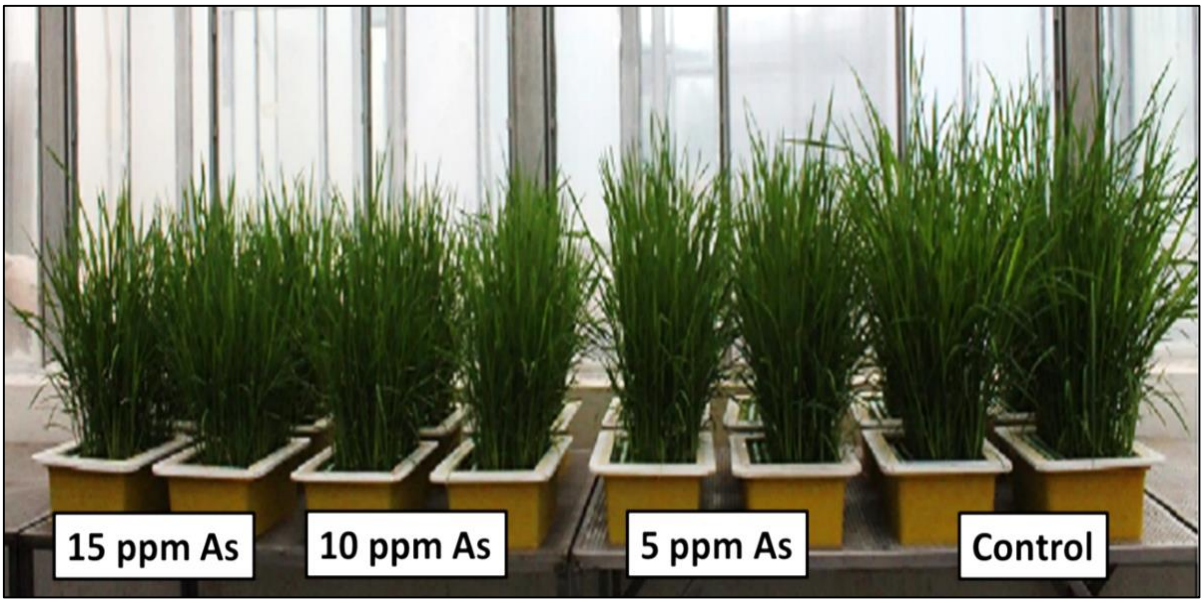


Figure 2.19. The screening system for As toxicity stress.
Plants are allowed to experience As stress for 18 days.

Table 2.7: Descriptive statistics and ANOVA results for different phenotypes.

	Trait	Chlorophyll content (SPAD)	Plant height (cm)	Root length (cm)	Shoot dry weight (g)	Root dry weight (g)	Arsenic content shoots (mg kg^{-1})	Arsenic content roots (mg kg^{-1})
Control	Min	33.02	38.09	9.05	2.01	1.40	ND	ND
	Max	37.98	71.83	17.99	3.99	2.49	ND	ND
	Mean	35.87	55.22	13.78	3.19	1.96	ND	ND
5 ppm As treatment	Min	30.03	26.26	7.00	1.02	0.90	5.32	78.06
	Max	35.99	64.96	15.99	3.50	1.90	13.83	110.84
	Mean	33.33	46.54	11.76	2.36	1.39	9.94	94.49
10 ppm As treatment	Min	27.02	24.26	6.02	0.92	0.51	8.38	109.83
	Max	30.99	51.92	11.99	2.79	1.50	17.82	150.76
	Mean	29.28	38.50	9.45	1.74	1.00	13.48	129.77
15 ppm As treatment	Min	23.01	16.04	4.05	0.60	0.14	13.57	116.46
	Max	28.98	38.92	9.99	1.89	1.07	25.71	165.90
	Mean	26.36	27.88	7.33	1.26	0.61	20.01	141.17
ANOVA result	G	***	***	***	***	***	***	***
	T	***	***	***	***	***	NA	NA
	G*T	***	***	***	***	***	NA	NA

Significance levels are indicated at P <0.001/*** Treatment and genotype treatment interaction effects were not analysed for As content traits since in control conditions arsenic is not measured. ND- not determined, NA- not applicable, G- genotype, T- treatment. SPAD- soil plant analysis development, mg kg^{-1}- milligram per kilogram, cm- centimetre, g- gram, As- arsenic, ppm- parts per million and ANOVA- Analysis of variance.

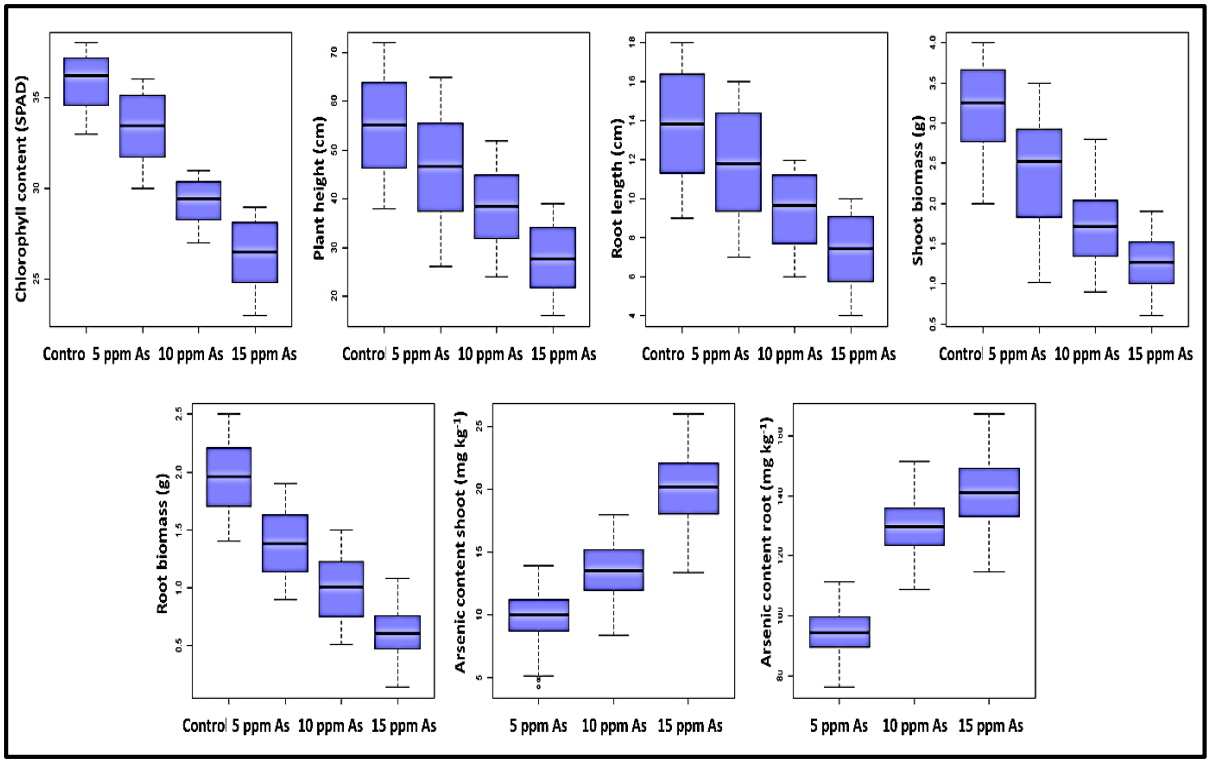


Figure 2.20. Box plot showing the distribution of phenotype values in 210 lines under arsenic stress for 18 days.

The thick line in the middle of the box is the median of the distribution while the lower and upper boundaries represent first and third quartile, respectively. Lower and upper whiskers are calculated based on 1.5 times the inter-quartile range. Control- without arsenic treatment, (5, 10 and 15 ppm) represents different levels of arsenic treatments (n=3).

Table 2.8: Summary of As content traits from field experiment 2015.

Trait		Arsenic content grain (mg kg^{-1})	Arsenic content husk (mg kg^{-1})	Arsenic content straw (mg kg^{-1})
Wet season 2015	Min	0.11	0.21	1.41
	Max	0.24	0.48	3.80
	Mean	0.18	0.34	2.54
Dry season 2015	Min	0.13	0.37	1.22
	Max	0.40	0.77	5.58
	Mean	0.26	0.57	3.46
Difference among seasons	t-test value	25.54	25.09	9.72
	p-value	0.00001*	0.00001*	0.00001*

Significance levels are indicated at P <0.001/*

2.7.3. As accumulation in mapping population under field condition

To evaluate the affinity of As accumulation in the selected breeding populations. A field screening experiment was set up in IRRI field station (the Philippines) during the dry and wet season of the year 2015. The soil in IRRI field station has an acceptable level of total As in the soil <7 mg kg^{-1} and the level of <0.02 mg l^{-1} in standard irrigation water. Samples harvested from the dry season show more accumulation in As than the samples harvested from the wet season. For both the seasons As accumulation was higher in the straw followed by husk than grain and follows the trend (straw>husk>unpolished grain) (Supplementary data 8). There was a significant difference in the mapping population for As accumulation in the population between the seasons (Figure 2.20). In wet season As content in grain ranged from 0.11 to 0.18 mg kg^{-1} and in dry season 0.13 to 0.40 mg kg^{-1} season (Figure 2.8).

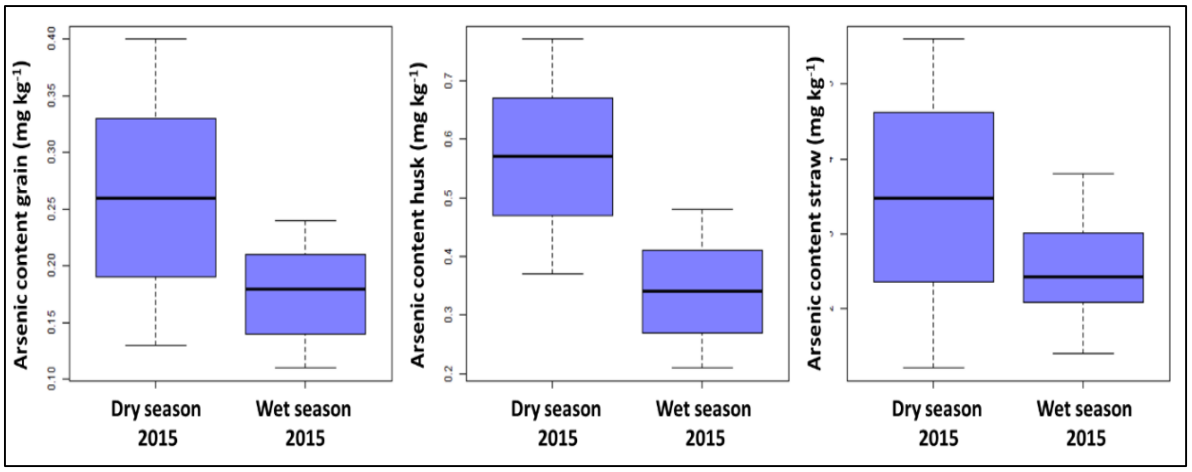


Figure 2.21. Boxplot showing phenotypic variation in the mapping population for As content traits from field experiment during dry/wet season 2015.

The thick line in the middle of the box is the median of the distribution while the lower and upper boundaries represent first and third quartile, respectively. Lower and upper whiskers are calculated based on 1.5 times the inter-quartile range. Control- without arsenic treatment, (5, 10 and 15 ppm) represents different levels of arsenic treatments (n=3).

2.7.4. Joint genome map of interconnected backcross breeding population

As the selected four population shares common recipient parent, SNPs generated from tGBS genotyping approach were combined to a single map based on the physical position of the SNPs (Figure 2.24). In the combined physical map, 11,515 polymorphic SNP sights remained after filtering out the SNPs having a high missing rate. These markers were, unevenly distributed across the genome, ranging from 608 SNPs on chromosome 9 to 1388 SNPs on chromosome 4 and with an average spacing of 31.8 kb between the SNPs. Detail description of SNPs for the combined population was summarised in Table 2.9

Table 2.9: Summary of SNPs in the combined population.

Chr.	LMD50 SNPs	Average distance(Kb)	Genome size(Kb)
Chr01	1111	38.2	42492.4
Chr02	1037	34.1	35401.9
Chr03	932	38.4	35824.4
Chr04	1388	24.4	33864.4
Chr05	1080	26.9	29100.3
Chr06	875	35.2	30809.5
Chr07	756	38.3	28942.5
Chr08	1014	27.4	27809.9
Chr09	608	35.1	21348.9
Chr10	868	22.6	19635.6
Chr11	1034	27.4	28312.7
Chr12	812	33.3	27023.4
Total	11515	31.8	360565.8

2.7.5. Population structure analysis

In order to analyse the presence of genetic diversity in the combined breeding population, structure analysis was carried out with the generated 11,515 polymorphic SNPs. Both the heat map of kinship relatedness matrix and the principal component analysis showed that two distinct and two-mixed population existed in the panel (Figure 2.2 A). The success of the backcross breeding with high intensity of selection was evident from the PCA analysis, as the developed lines were genetically closer towards the recurrent parent WTR1. A small number of lines from the cross WTR1 × OM997 and WTR1 × BG 300 were classified into WTR1 × Zhong 413 population (Figure 2.2 B). Population structure analysis based on Bayesian approach also

revealed the existence of two distinct population and two-admixture population exist in the selected breeding population (Figure 2.22).

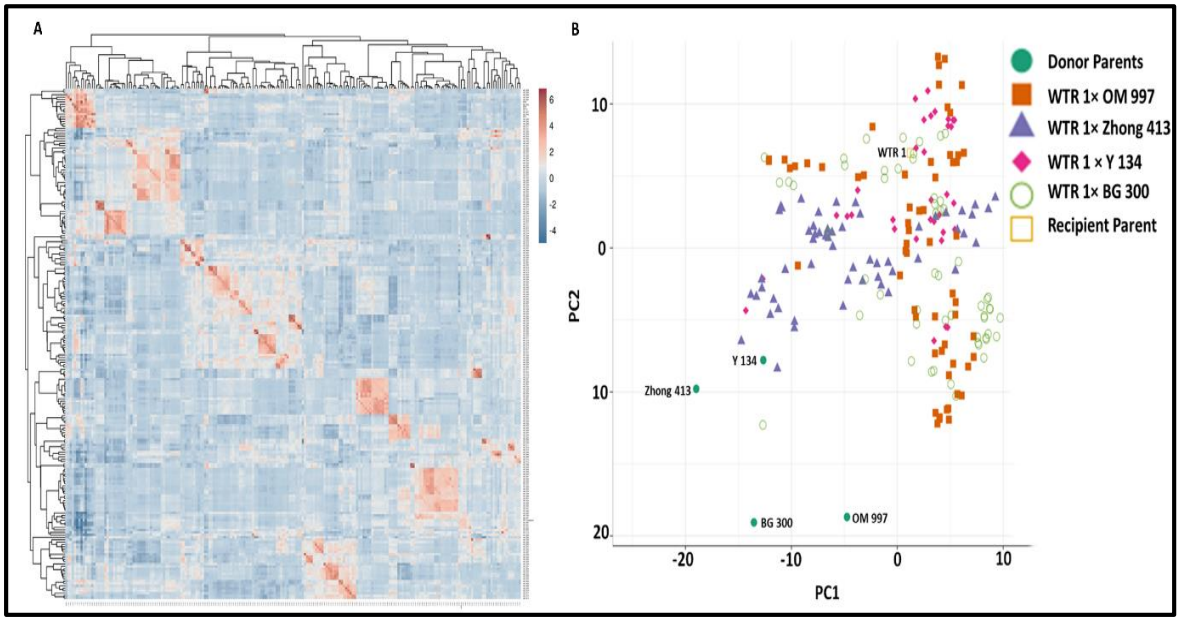


Figure 2.22. Population structure analysis.

(A) Heat map of kinship with the tree shown on the top and left. (B) The plot of the first two principal components in the selected populations.

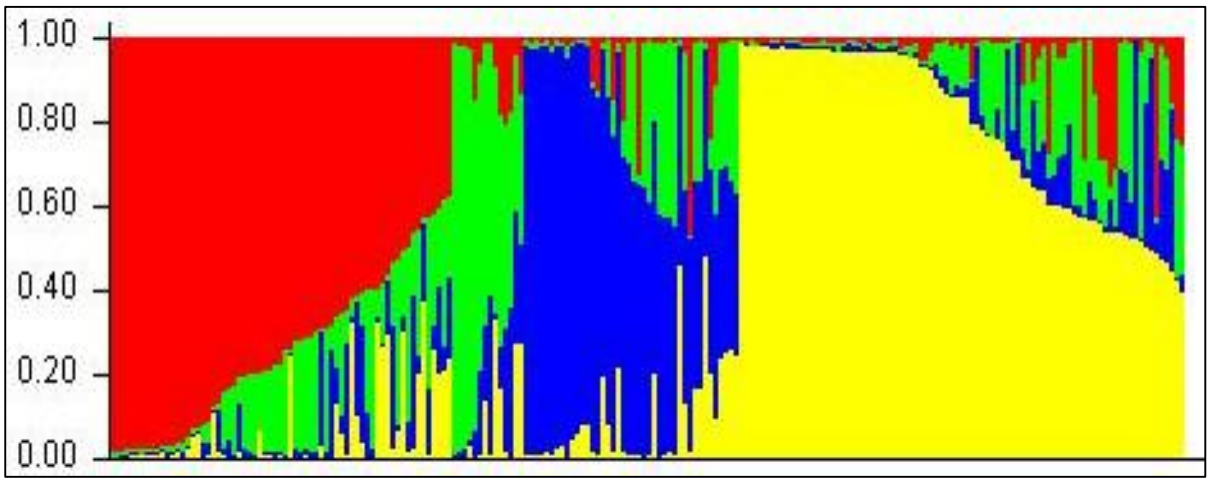


Figure 2.23. The pattern of variation in 210 lines based on 11,515 SNP markers.

Bar length and different colour, represent the membership probability of accessions belonging to different subgroups. Run with the highest likelihood ratio of K value of 4 was used to define the population structure.

2.7.6. QTLs associated with As related traits

In total 43 SNPs showed association revealed significant marker-trait association ($P>1.0\times10^{-4}$) for the As-related traits (Figure 2.23). Significant markers with a linkage disequilibrium value ($R^2 > 0.87$) were considered as a putative QTL interval and 19 QTLs were defined by keeping the associated SNPs (Table 2.11). Among the 19 QTLs, two QTLs were identified for As content in the grain for dry season 2015. In 5 ppm hydroponics As treatment condition, 9 QTLs were identified for As content in the shoot, 7 QTLs for relative plant height and one QTL for relative shoot biomass. QTLs for As content in the grain *qGrAsC*6 and *qGrAsC*12 were mapped on chromosome 6 and 12 respectively. As content shoot QTLs *qShAsC*1, *qShAsC*5, *qShAsC*5.1, qShAsC5.2, *qShAsC8, qShAsC*8.1, *qShAsC*11, *qShAsC*11.1 and *qShAsC*11.2 were mapped on 1, 5, 8 and 11 respectively. For relative plant height QTL *qRph*1, *qRph*2, *qRph*5, *qRph*5.1, *qRph8, qRph*8.1 and *qRph*11 were mapped on 1, 2, 5, 8 and 11 respectively. For Relative shoot biomass, one QTL *qRshbms*3 was mapped on chromosome 3. Excitingly, all the QTLs except *qRph*2, all the QTLs identified with relative plant height was co-localised with the QTLs identified for the As shoot content. In 10 ppm As treatment, all the QTLs identified under 5 ppm As treatment except relative shoot biomass QTL *qRshbms*3 was identified again in the 10 ppm As treatment. In 15 ppm As treatment, Relative plant height QTLs identified in 5 ppm and 10 ppm As treatment (*qRph*1, *qRph*2, *qRph*5, *qRph*5.1, *qRph8, qRph*8.1 and *qRph*11) was again for a third time detected with different *P*-value. Colocalization of QTLs identified in varying concentration of As treatments confirm the accurateness of the identified QTLs. All the QTLs identified for the As related traits were condensed in the table (Table 2.10).

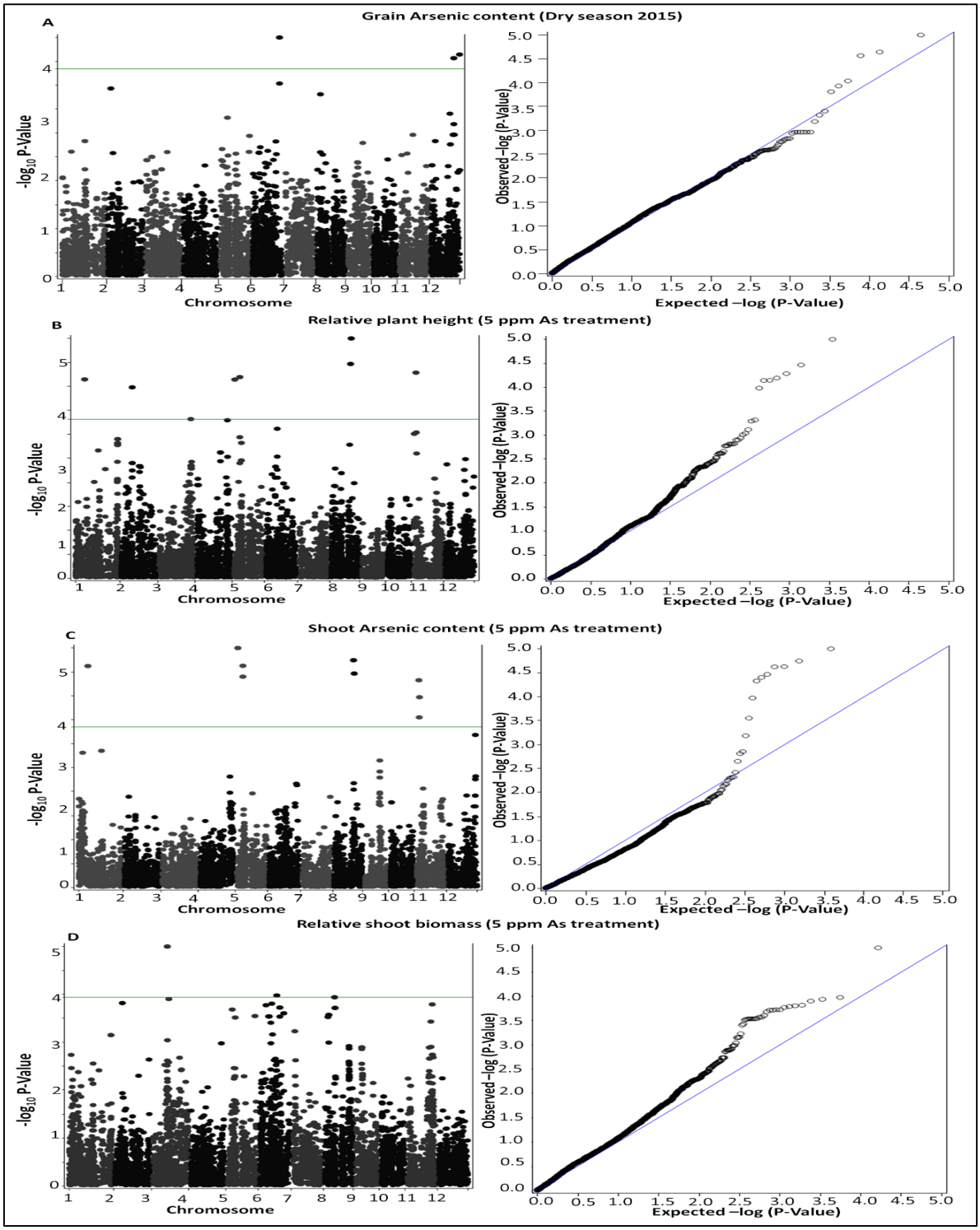


Figure 2.24. Genome-wide association mapping result.

Manhattan plot of association mapping results with p-value on (left) and corresponding Quintile- Quintile plots are shown at (right). The analysis was performed for the entire population. The green line in the plot indicates threshold value ($P>1.0\times10^{-4}$) and was adopted from previous study.

Table 2.10: Summary of quantitative trait loci identified for As related traits using genome-wide association analysis

Environment	Trait	QTL	Chr.[a]	Interval(bp)[b]	Peak SNP[c]	Position (bp)	Allele[d]	MAF[e]	Effect[f]	*P*-value[g]
2015 Dry season	Grain As content	*qGrAsC6*	6	27560424-27935847	S6_27560457	27560457	C/T	0.25	-0.19	2.51E-05
		qGrAsC12	12	27318964-27444191	S12_27318964	27318964	T/C	0.30	0.04	5.35E-05
5 ppm As Treatment	As content shoot	*qShAsC1*	1	8992538-9145647	S1_9099652	9099652	A/G	0.16	-1.79	2.99E-06
		qShAsC5	5	27295-358908	S5_338431	338431	C/T	0.19	-1.35	1.06E-06
		qShAsC5.1	5	5489846-5519035	S5_5489846	5489846	A/G	0.28	-0.79	2.95E-06
		qShAsC5.2	5	5489846-5519035	S5_5489888	5489888	T/C	0.25	-1.86	5.51E-06
		qShAsC8	8	19375200-19396300	S8_19391086	19391086	A/G	0.36	-1.32	2.14E-06
		qShAsC8.1	8	20024757-20039777	S8_20039777	20039777	C/T	0.28	-2.11	4.61E-06
		qShAsC11	11	1868510-2004854	S11_2004825	2004825	G/A	0.29	-1.60	6.76E-06
		qShAsC11.1	11	2325462-2476215	S11_2347253	2347253	G/A	0.20	-1.75	5.74E-05
		qShAsC11.2	11	2564688-2632512	S11_2632503	2632503	C/T	0.29	-1.61	1.80E-05
	Relative plant height	*qRph1*	1	8992538-9145647	S1_9099652	9099652	A/G	0.16	-1.79	9.85E-06
		qRph2	2	10768489-10924463	S2_10922783	10922783	C/T	0.18	-2.19	1.56E-05
		qRph3	3	32737656-32799136	S3_32737656	32737656	T/C	0.14	-0.37	9.92E-05
		qRph5	5	27295-358908	S5_338431	338431	C/T	0.19	-1.35	9.95E-06
		qRph5.1	5	5489846-5519035	S5_5489888	5489888	T/C	0.25	-1.86	8.62E-06
		qRph8	8	19375200-19396300	S8_19391086	19391086	A/G	0.36	-1.32	4.01E-06
		qRph8.1	8	20024757-20039777	S8_20039777	20039777	C/T	0.28	-2.11	9.09E-07
		qRph11	11	1868510-2004854	S11_2004825	2004825	G/A	0.29	-1.60	6.66E-06
	Relative shoot biomas	*qRShBms3*	3	14738851-14958669	S3_14936266	14936266	G/C	0.19	-1.59	8.44E-06
10 ppm As Treatment	As content shoot	*qShAsC1*	1	8992538-9145647	S1_9099652	9099652	A/G	0.16	-1.79	2.39E-06
		qShAsC5	5	5489846-5519035	S5_5489888	5489888	T/C	0.25	-1.86	5.61E-08
		qShAsC5.1	5	27295-358908	S5_338431	338431	C/T	0.19	-1.35	9.40E-07
		qShAsC5.2	5	5489846-5519035	S5_5489846	5489846	A/G	0.28	-0.79	1.90E-06
		qShAsC8	8	19375200-19396300	S8_19391086	19391086	A/G	0.36	-1.32	3.32E-07
		qShAsC8.1	8	20024757-20039777	S8_20039777	20039777	C/T	0.28	-2.11	5.32E-07
		qShAsC11	11	1868510-2004854	S11_2004825	2004825	G/A	0.29	-1.60	1.08E-05
		qShAsC11.1	11	2325462-2476215	S11_2347253	2347253	G/A	0.20	-1.75	4.96E-05
		qShAsC11.2	11	2564688-2632512	S11_2632503	2632503	C/T	0.29	-1.61	2.67E-05
	Relative plant height	*qRph1*	1	8992538-9145647	S1_9099652	9099652	A/G	0.16	-1.79	4.81E-05
		qRph2	2	10768489-10924463	S2_10922783	10922783	C/T	0.18	-2.19	2.90E-06
		qRph3	3	32737656-32799136	S3_32737656	32737656	T/C	0.14	-0.37	5.08E-05
		qRph5	5	27295-358908	S5_338431	338431	C/T	0.19	-1.35	9.17E-05
		qRph5.1	5	5489846-5519035	S5_5489888	5489888	T/C	0.25	-1.86	4.81E-05
		qRph8	8	19375200-19396300	S8_19391086	19391086	A/G	0.36	-1.32	5.05E-05
		qRph8.1	8	20024757-20039777	S8_20039777	20039777	C/T	0.28	-2.11	7.39E-06
		qRph11	11	1868510-2004854	S11_2004825	2004825	G/A	0.29	-1.60	2.72E-05
15 ppm As Treatment	Relative plant height	*qRph1*	1	8992538-9145647	S1_9099652	9099652	A/G	0.16	-1.79	1.55E-05
		qRph5	5	27295-358908	S5_338431	338431	C/T	0.19	-1.35	1.15E-05
		qRph5.1	5	5489846-5519035	S5_5489888	5489888	T/C	0.25	-1.86	1.23E-05
		qRph8	8	19375200-19396300	S8_19391086	19391086	A/G	0.36	-1.32	5.40E-06
		qRph8.1	8	20024757-20039777	S8_20039777	20039777	C/T	0.28	-2.11	2.11E-06
		qRph11	11	1868510-2004854	S11_2004825	2004825	G/A	0.29	-1.60	9.50E-06

The threshold for significance, *P*-value <1.00× 10^{-4}. [a] Chr: Chromosome, [b] Defined based on LD block, [c] Significant SNP with the best *p*-value, [d] Allele (major/minor alleles), [e] MAF: minor allele frequency, [f] Effect: allele effect concerning the minor allele, [g] *P*-value for the significant SNP.

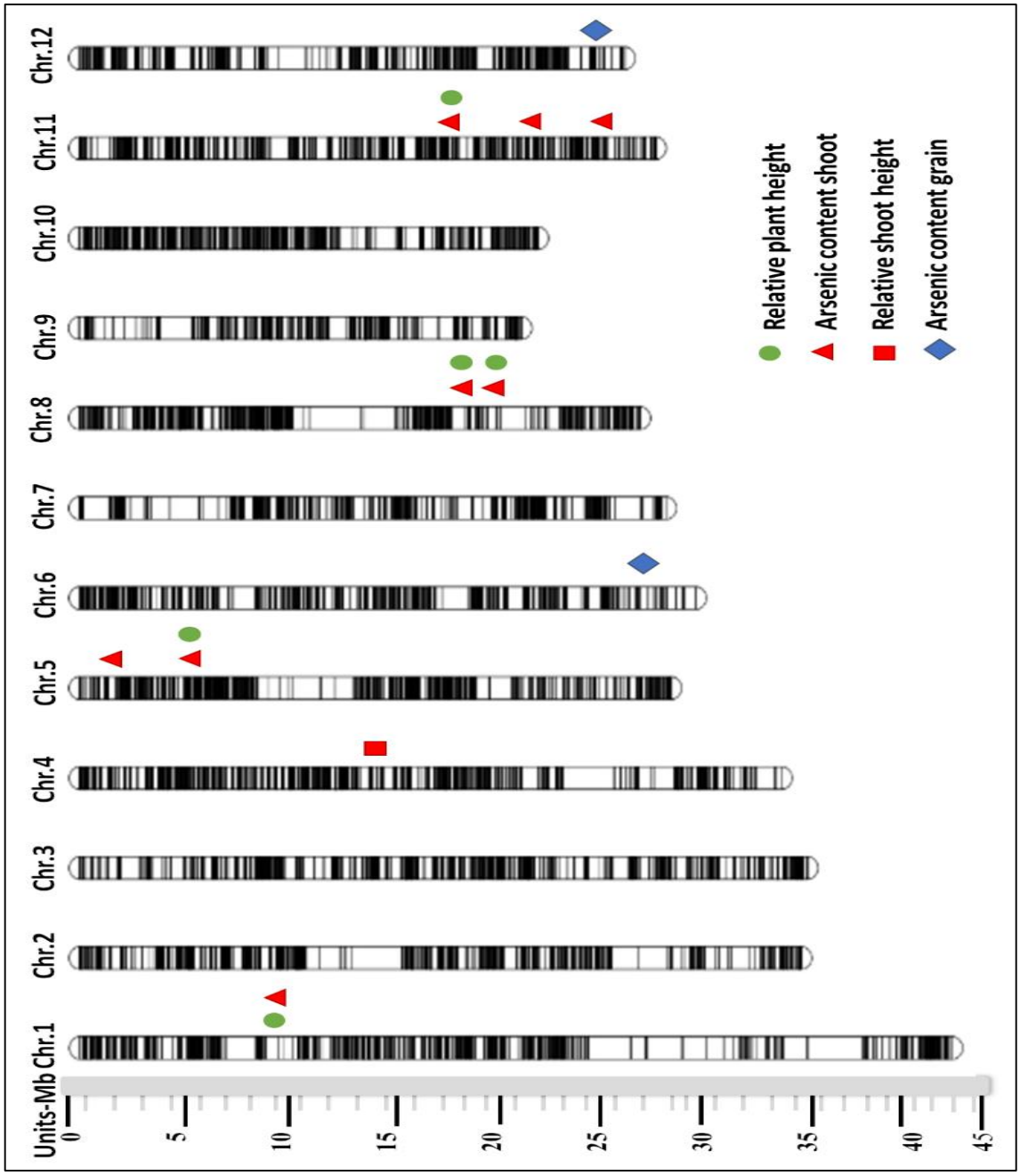


Figure 2.25. Chromosomal distribution of LMD50 polymorphic single nucleotide polymorphisms (SNPs) of the combined population and detected quantitative trait loci (QTLs).

QTLs were located on the chromosome based on the physical position of the SNP marker. Chr- chromosome, Mb- millions of base pairs

2.7.7. Selection of candidate genes associated with As stress

In the LD interval of the identified QTLs, 153 gene models were present in the MSU Rice Genome Annotation Database. In the identified 153 genes, 810, 939, 1181, 994 and 1221 polymorphic SNPs were identified between the parents OM997, BG300, Y134, ZHONG413 and WTR1 respectively in the Rice SNP-Seek Database (Supplementary data 9). Among those identified polymorphic SNPs, 7.2% of SNPs showed non-synonymous mutations in 76 genes (Table 2.11). Expression data of these genes were retrieved from a study by (Yu et al. 2012) using the web application Rice Expression Database (Xia et al. 2017). Thirty-one genes out of the 76 most likely candidate genes showed expression responses to the sodium Arsenite (As^{III}) stress treatment according to expression data (Yu et al. 2012). Among 30 genes showing expression responses, seven genes were upregulated, and 29 genes show significant down-regulation (Figure 2.25).

Table 2.11: Useful SNPs identified in the QTL regions using 3K RGP datasets.

QTLs*	Chr[a].	Polymorphic SNPs[b]					Non-synonymous SNPs[c]					Genes[d]	
		OM997	BG300	Y-134	ZHONG 413	WTR1	OM997	BG300	Y-134	ZHONG 413	WTR1	Total	Non-synonymous
qGrAsC6	6	8	33	91	27	8	1	4	9	4	2	30	5
qGrAsC12	12	76	74	80	78	78	8	7	9	8	8	17	6
qShAsC1, *qRph1*	1	111	51	4	2	2	5	5	0	0	0	13	5
qShAsC5, *qRph5*	5	172	267	330	320	323	19	40	51	51	52	36	19
qShAsC5.1, *qShAsC5.2*, *qRph5.1*	5	0	15	14	0	14	0	4	5	0	4	2	2
qShAsC8, *qRph8*	8	8	8	8	9	9	0	0	0	1	1	1	1
qShAsC8.1, *qRph8.1*	8	18	19	18	17	20	1	0	1	0	4	1	1
qShAsC11, *qRph11*	11	171	179	177	176	184	34	38	38	34	38	16	12
qShAsC11.1	11	162	143	143	163	146	19	18	16	19	16	13	9
qShAsC11.2	11	54	90	53	10	146	4	7	5	0	11	6	4
qRph2	2	4	34	238	165	263	1	5	53	30	58	10	9
qRph3	3	26	26	25	27	28	5	5	4	5	5	8	3
Total		**810**	**939**	**1181**	**994**	**1221**	**97**	**133**	**191**	**152**	**199**	**153**	**76**

*QTLs identified from field and hydroponics experiment, a-chromosome, b- total number of polymorphic SNPs present in the QTL interval, c- SNPs showing a mutation in the coding region of gene and d- genes present in the QTL interval.

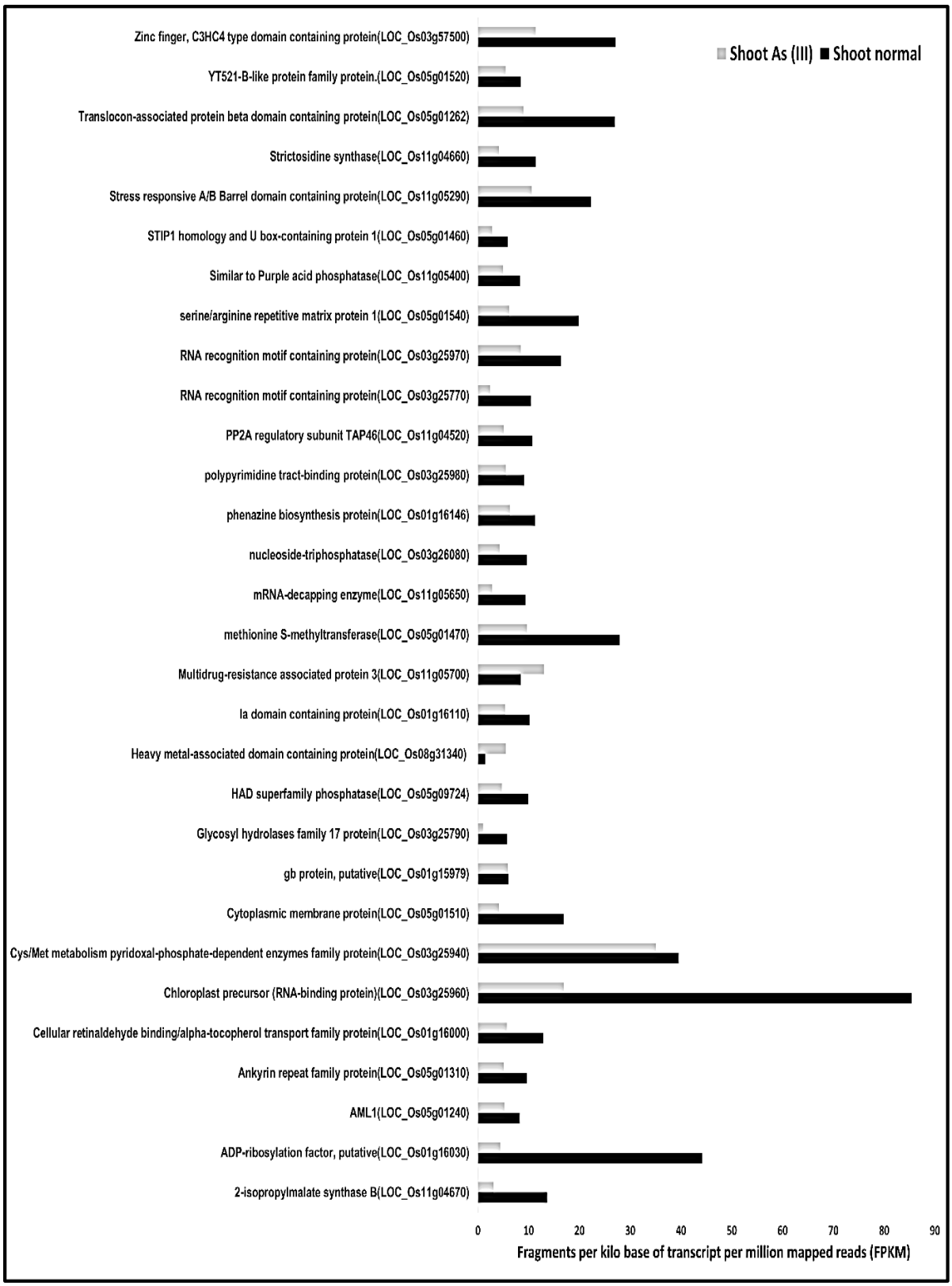


Figure 2.26. Transcriptional regulation of genes contained in mapped QTL regions. Only genes containing non-synonymous SNPs showing significant expression response under $As^{(III)}$ stress are plotted. Gene expression data adopted from Yu et al. (2012)

3. Discussion

3.1. Arsenic contamination in rice

Rapidly growing populations need more food despite the increased labour cost, reduced farmland, and water scarcity that threaten rice production. Long-term use of As-contaminated groundwater for irrigating crops has resulted in a significant increase in As in the topsoil, thereby contaminating the food chain (Pandey et al. 2015). Governments of developing rice growing Asian nations are encouraging farmers to adopt modern technologies such as direct-seeded rice (DSR) and alternate wetting-drying (AWD) as a cost-efficient sustainable method to reduce the environmental footprint of rice production (Richards and Sander 2014; Singh Chauhan et al. 2015). Toxic levels of As in the topsoil can potentially affect the performance of DSR by hindering the most delicate rice germination process and early seed establishment of the rice growth cycle (Abedin and Meharg 2002; Shri et al. 2009). Water management using AWD was proposed as one of the strategies to control As bioavailability in the soil-plant system (Mitra et al. 2017). However, rice genotypes suitable for the DSR and AWD system under As-toxic soil have not been identified and tested (Suriyagoda et al. 2018). In our current study, we attempt to reveal the variations among the genotypes for As tolerance in the germination and seedling stage. Also, variation in genotypes for As accumulation in grain. Results obtainable from our study will benefit rice breeders to select suitable genotypes for development of As-safe varieties suitable for modern rice production technologies (Tuli et al. 2010).

3.2. Arsenic tolerance in rice breeding panel

Germination assays are not a commonly used technique for testing As toxicity in rice. However, in As-contaminated soils, seed germination capability is one of the best indicators for successful or unsuccessful development of a plant (Abedin and Meharg 2002). Presented results revealed the toxic effect of As on seed germination ability. Seed germination ability decreased significantly with increasing As concentration in the treatment, which is consistent with the results of the previous study (Abedin et al. 2002b; Halim et al. 2016). Arsenic is known to inhibit overall growth and development of rice plant (Azizur Rahman et al. 2007; Shri et al. 2009). In hydroponics As stress, irrespective of the rice genotype the overall performance of the plant was decreased with the increasing concentration of As in the nutrient solution (Figure 2.2). As burden was relatively more in the root than the shoot (Root > Shoot) and in

acceptance with the results of the prior study (Raab et al. 2007; Lu et al. 2010). However, considerable genotype variation is observed among the studied genotypes for As tolerance and accumulation in the shoot (Figure 2.3). In general, from the hydroponic screening experiment, it is noted that genotypes which are including As in the shoot displayed decrease As tolerance. In a prior study, involving mega variety IR64, it was observed that IR64 was susceptible to $As^{(V)}$ treatment with high accumulation in the shoot (Tripathi et al. 2012). In the present experiment, IR64 display tolerance towards As by including in the shoot and correlates with the results of a previous study (Dasgupta et al. 2004). Astoundingly, IR64 from field condition was one of the genotypes, which has less As in the unpolished grain (< 0.20 mg kg^{-1}). Genotypic differences in As accumulation by rice genotypes under the same soil environmental conditions is necessary to reveal the genotypic difference (Kuramata et al. 2011; Islam et al. 2016). Total As concentrations in the grain from the 53 genotypes (Figure 2.4) found that As in unpolished grain (brown rice) ranged from 0.12 to 0.48 mg kg^{-1} with an average value of 0.31 mg kg^{-1}. Most of the studied genotypes surpass the allowed maximum level of inorganic As 0.20 mg kg^{-1} set by Codex Alimentarius Commission (United Nations food safety standards) (Pillai et al. 2010; Stankovic 2015; Islam et al. 2017). It is also evident from the present experiment that among the tested genotypes (BR11 and BR28) from Bangladesh accumulate more As in the grain (>0.40 mg kg^{-1}) and display strong susceptibility in germination and hydroponic condition by including As in the shoot. Similar results were observed with the same genotypes in the field based experiments (Azizur Rahman et al. 2007; Ahmed et al. 2011). Arsenic accumulation in unpolished grain pattern reveals that most of the *O.Japonica* subgroup genotypes trend to accumulate more As in the grain than *O.Indica* subgroup (Tuli et al. 2010). However, Haoannong an *O.Japonica* genotype from south China accumulate less As <0.12 mg kg^{-1} in grain and also displayed high resistance in germination tolerance but displays susceptibility in hydroponic screening by excluding As in the shoot. In our current study, we attempt to explore the genotypes commonly used by breeders to develop varieties for Bangladesh and India. In developing rice growing Asian countries, straw was either burned in the field for getting rid it or used as fodders for cattle's, which further increase the risk of As exposure (Njie and Reed 1995; Lawler and Dritz 2005). Arsenic tolerant or susceptible genotypes with As excluders and less accumulation in grain are the most desirable characteristics in genotypes for breeding As-safe varieties.

3.3. Development of a permanent breeding population

In plant breeding, the detection of quantitative trait loci (QTL) is no longer limited by the availability of genetic marker information and genotyping throughput, but rather by choice of the genetic material used. QTL mapping is often performed in crossing experiments where plants are randomly selected from a pool of all potential progeny (McCough and Doerge 1995). For instance, the population development using the MAGIC design (Multiparent advanced generation intercross) involving intercrossing several times with multiple founders to combine the genetic material of all the founder in a single population are a clear advantage in mapping QTLs compared to a classical bi-parental population (Bandillo et al. 2013; Raghavan et al. 2019). However, the MAGIC population is developed through random matting design, QTLs detected from such population are vital for us to understand the genetic mechanisms governing a complex trait, but may not be directly relevant to plant breeding as they are not detected from the breeding population (Zhang and Gai 2009; Cui et al. 2015).

Using permanent breeding population (population derived from the targeted breeding program) for mapping QTLs serve a dual purpose, permanent mapping populations for precise QTL mapping and direct use in variety development (Bandillo et al. 2013). However, development of such population takes a very long time, and success depends on the objective of the breeding program. Unlike a set of naturally diverse germplasm, the two developed breeding population is tailor-made with a combination of useful traits derived from the eleven elite donor cultivars and stabilised for yield-related traits (Ali et al. 2018b). The developed breeding populations also present opportunities for studying the interactions of genome introgression and chromosomal recombination. However, selected breeding populations often have depleted genetic variation with small population sizes, resulting in low power in detecting useful QTLs (Cui et al. 2015). In order to overcome the small size of the breeding population, all the developed breeding population in this study shared a common recipient parent WTR1 were combined to a single mapping population. It was proposed to detect As related QTLs in selected breeding populations by either using a single backcross breeding population or by combining multiple backcross breeding populations.

3.4. Advanced genotyping in the interconnected breeding population

3.4.1. Advantages in tGBS genotyping over convention GBS

Cost effective next-generation sequencing has been successfully employed for whole

genome sequencing, gene expression, and single-nucleotide polymorphism (SNP) discovery (Xu et al. 2011; Harper et al. 2012; 2014; Li et al. 2014). Several approaches and methods are already developed for SNP discovery and genotyping in several crop species (Elshire et al. 2011; Wang et al. 2012). Genotyping-by-sequencing (GBS) has also emerged as a powerful breeding tool with the continued increase of sequencing output, the development of reference genomes, and improved bioinformatics. Plant scientists are getting deep into connecting phenotype to genotype using sequencing outputs. Deciphering interactions among heritable genetic factors and phenotypes will aid in harnessing the benefits of genomics-assisted selection in plant breeding. GBS has the potential to discover novel or population-specific polymorphisms. One approach to using GBS is to incorporate polymorphisms discovered via GBS into a closed platform, like an array, which can then be used to genotype an entire population of interest (Poland and Rife 2012). These array based, large- scale SNP discovery pipelines have advanced considerably in the identification of chromosome-specific SNPs involved in stress tolerance mechanisms (Akpinar et al. 2017; Lucas et al. 2017). Conventional GBS (cGBS) (Elshire et al. 2011; Poland and Rife 2012; Poland et al. 2012; Li et al. 2015) involves multiplexing samples using DNA barcoded adapters and a reduction in genome complexity (using, for example, restriction endonucleases to target only a small portion of the genome) to produce high-quality polymorphism data at a relatively much lower cost per sample. This approach has been demonstrated to be robust across a range of species and to be capable of producing enormous amounts of molecular markers (Elshire et al. 2011; Poland et al. 2012; Li et al. 2015). However, cGBS comes with some complexities such as yielding relatively fewer reads per site and having high levels of missing data across samples. Low read depths per site limit the ability of cGBS to identify heterozygous loci in diversity panels while high levels of missing data require imputation that limits the ability of cGBS to detect rare alleles. Like many modified and improved versions of cGBS, the tunable genotyping-by-sequencing (tGBS®) technology (Schnable et al. 2013) overcomes these two challenges of cGBS by amplifying and, therefore, sequencing fewer sites. This result in a given number of sequence reads being distributed across fewer sites of the genome, thereby yielding more reads per site. This higher read depth per site results in less missing data from an individual to individual, which provides more repeatability and enhances the usability of the resulting genotyping data. Also, because of this higher read depth, it is possible to accurately call heterozygous loci

and confidently detect novel or rare alleles. There are two basic modifications in tGBS® relative to cGBS. Firstly, tGBS® uses barcoded single-stranded oligos instead of the double-stranded adapters used in cGBS, thereby eliminating the problem of the inter-molecular ligation of adapter molecules and making the entire process simpler and improving the quality of results. Secondly, the modification that overcomes the read depth problem associated with cGBS involves selective amplification of restriction fragments. The amount of genome reduction levels (GRLs) can be tuned. For example, in GRL1, GRL2, and GRL3, only ¼, 1/16th, and 1/64th of all of the restriction fragments are sequenced. To achieve these levels of genome reduction, about one to three additional nucleotides are added up at the end of a selective polymerase chain reaction (PCR) primer. These GRLs concentrate the available sequencing reads at fewer sites of the genome, thereby increasing read depth and enhancing the accuracy of SNP calling, including in heterozygous individuals, as well as the discovery of rare alleles (Patrick S. Schnable; Schnable et al. 2013; Islam et al. 2015). The use of multiple parents in the development of permanent lines through backcrosses is advantageous. In genetic mapping, the backcross population harnesses more allelic diversity thereby enabling the detection of QTLs with more precision as compared to using bi-parental populations (Zhu et al. 2015; Ali et al. 2017). Previously, tGBS® has been shown to have several advantages. In a study involving upland cotton (*Gossypium hirsutum* L.) (Islam et al. 2015), the tGBS® protocol (Schnable et al. 2013) yielded more high-quality SNPs with higher read depths per SNP site than GBS (Elshire et al. 2011). This frequency of SNPs in the coding region is higher than that reported in earlier studies (Arai-Kichise 2011; Subbaiyan 2012; Jain et al. 2014; Mehra et al. 2015). Four % of SNPs were detected in regulatory regions. The variation in the distribution of polymorphisms on chromosomal basis has frequently been reported in rice and model plants (Feltus et al. 2004; Nordborg et al. 2005; Arai-Kichise 2011; Hu 2014). This localised effect of the chromosomal distribution of SNPs is attributed to GBS technologies which rely on uniquely aligned reads resulting in a non-uniform distribution of unique sequences (Elshire et al. 2011; Schnable et al. 2013) and natural selection-sweeps during rice domestication (Caicedo 2007). It is notable that the introgression patterns do not appear to be random, providing suggestive evidence of selection during backcrossing and in subsequent breeding generations.

3.4.2. The functional effect in SNPs detected by tGBS genotyping

Most of the 11 sub-populations exhibited similar introgression patterns, noticeably corresponding to the breeding strategy for their development. Among the 4,784 non-synonymous SNPs, 426 were predicted based on SIFT analysis to confer deleterious effect on gene function and were predicted to be highly detrimental, having a tolerance index of 0.00. In all the 11 sub-populations, 102 loci contained 120 deleterious large-effect SNPs, with 1–4 SNPs per loci. Of the 102 affected loci, 24 were predicted to be responsive to biotic (6 loci) and abiotic (18 loci) stress. These loci contained deleterious SNPs that substitute the amino acid and change the function of protein either positively or negatively. The identified abiotic and biotic stress-responsive loci are represented in Circos diagrams (Figure 3.1 and Supplementary figure 3.2 and 3.3). In sub-populations 1 (*Dp: Haoannong*) and 4 (*DP: Y-134*), a G/A deleterious SNP at position 349360 in the *Os01g01689* locus alters an amino acid from *Ala* to *Thr* (Supplementary data 10). This locus is associated with the abiotic stimulus. The DP *Haoannong* of sub-population one is known for its drought and salinity tolerance (Li and Xu 2007) while the *Y-134* donor is suitable for agronomic traits – characteristics that were used as the basis for their selection as DPs in the early-backcross introgression-breeding program. A large-effect deleterious allele was observed in sub-populations 9 (*DP: Basmati*), 10 (*DP: M-401*), and 11 (*DP: X-21*) at position 6884254 in the *Os11g12330* locus and the changed codon A/C changes the amino acid from *Lys* to *Asn* (Supplementary data 8). The presence of this SNP was observed in the DPs while there was no variation in the RP at this position, suggesting that this allele was introgressed from the DPs in the sub-populations. High variability was observed at position 16687362 of the *Os12g28250* locus within sub-populations 7 (*DP: BG-300*), 9 (*DP: Basmati*), and 10 (*DP: M-401*) as these three DPs have “G” at this position, the same as in the reference genome (MSU7), while the RP has SNP “C” (Supplementary data 10). The presence of this allele precisely defines the background of the RP (WTR-1) in the sub-populations.

3.4.3. Breeding population panel

Through the novel early-backcross introgression-breeding strategy used in this study, many important tolerance alleles were combined due to selections made simultaneously in different stress conditions over three rounds using these sub-populations. The genetic distance between the recipient parent and the 564 lines can

be explained by the novel breeding strategy used in this study (Ali et al. 2006, 2012a, b, 2013, 2017). Each lines arising from the cross between each of the 11 DPs and a single recipient parent underwent rigorous screening and selection under various abiotic stresses. The whole panel of 564 genotypes possesses varying levels of tolerance towards these multiple stresses in the field screenings (unpublished data).

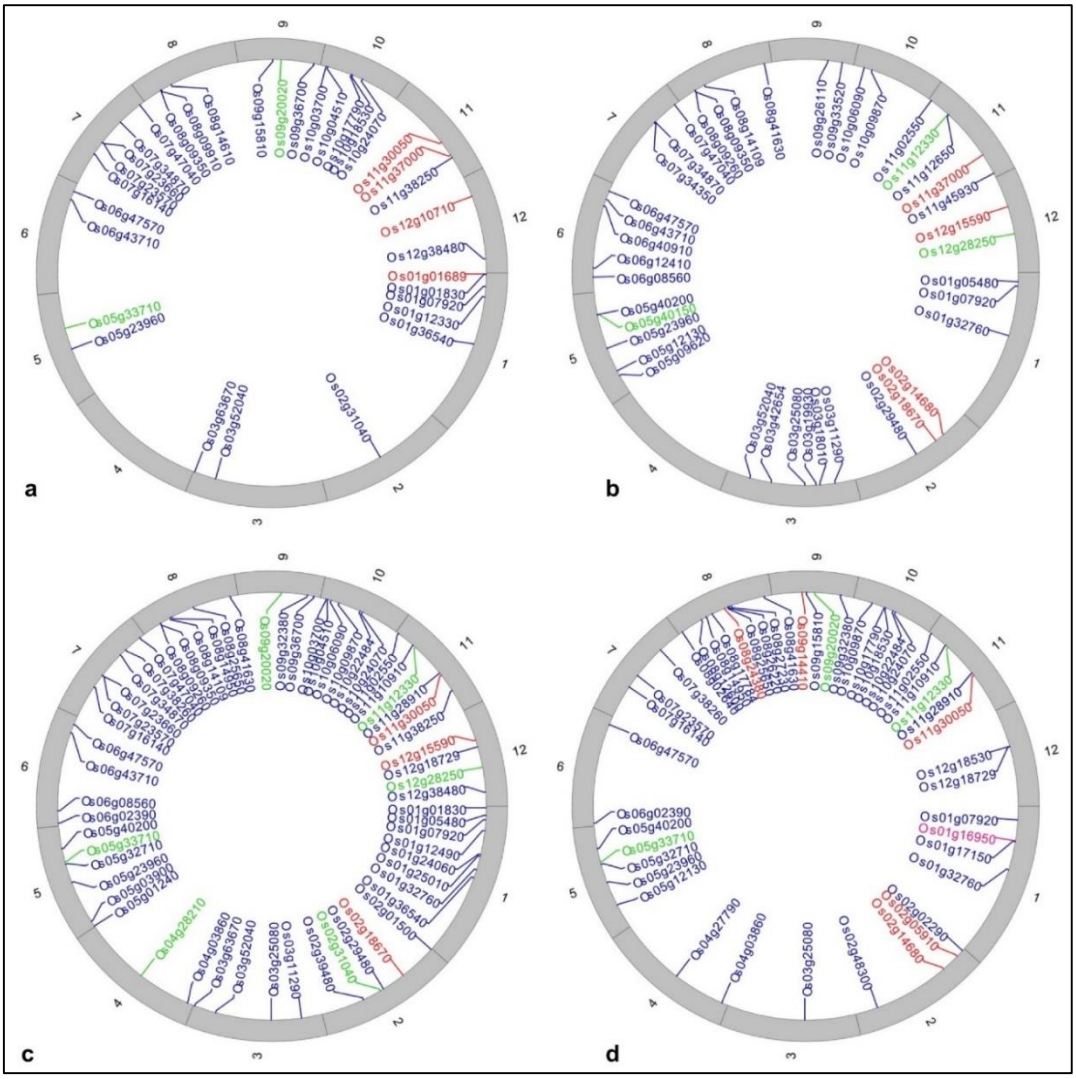


Figure 3.1. Distribution of non-synonymous SNPs that have a deleterious effect on proteins.

Circos diagrams representing annotated gene locations in (a) sub-population 1 (DP: *Haoannong*); (b) sub-population 9 (DP: *Basmati*); (c) sub-population 10 (DP: *M-401*); and (d) sub-population 11 (DP: *X-21*). The outer numbering shows the 12 rice chromosomes. Red colour shows the abiotic stress-related loci, green colour shows biotic stress-related loci, and blue colour shows the loci other than stress-related families.

The introgression event of single backcrossing and subsequent stringent screening and selection plausibly resulted in major functional introgression from the donor genomes and quick fixation of the following genetic effects through novel breeding strategy. The evident genomic introgression in each IL from the DP (Supplementary figure 3.4–3.6) is consistent with the hypothesis that the introgressed genomic regions contributed to the strong phenotypic responses in resulting breeding lines, rendering them tolerant to multiple biotic and abiotic stresses during stringent selection. The favourable phenotypic effects are stabilised in the population through positive selection leading to uniqueness and deviance from the recipient parent, both genotypically and phenotypically.

3.5. The arsenic stress responses in the breeding population

3.5.1. Single backcross breeding population

In this study, it was exposed two parental lines (WTR1/Haoannong) and 194 BC_1F_6 lines to a concentration of 10 ppm As) at the seedling stage for 18 days in a hydroponic-based experimental setup. Similar concentrations frequently occur in the topsoil of the rice-growing regions of Bangladesh and West Bengal of India (Abedin et al. 2002a; Zhang et al. 2008). When exposed to 10 ppm of As, a wide range of morpho-physiological changes occurred in the plants, leading to inhibited plant growth with a rapid decline in biomass and reduced chlorophyll content when compared to the control (Figure 1) (Wu et al. 2011; Pandey et al. 2015). In the current experiment, the distribution of As showed similar patterns (root>shoot), in agreement with the previously reported genotype-dependent As accumulation and distribution, as the results suggested that the *Indica* recipient parent (WTR1) accumulated less As in shoots than its *japonica* donor parent (Haoannong) (Rahman et al. 2007; Bhattacharya et al. 2010; Wu et al. 2011). Correlation among the traits revealed that plant performance under As stress depended on As concentration in the tissue, with increased uptake in shoots reducing the plant‘s overall performance. Arsenic translocation to shoots was positively correlated with As content in roots, which indicates that avoiding As translocation to shoot tissue will, in turn, increase plant performance by an As avoidance mechanism. In one earlier study, a doubled-haploid (DH) population derived from cross CJ06 × TN1 (*Japonica* × *Indica*) cultivars were used to evaluate the toxic effect of 10 ppm As in a soil-based controlled greenhouse

experiment. The correlation in that study revealed results similar to those of the current study, clearly indicating that the performance of plants under As stress is inversely related to the As content in shoots (Figure 3) (Zhang et al. 2008). Arsenic toxicity traits observed in the rice population in our experimental conditions signify that As tolerance and accumulation in rice seedlings is a quantitatively inherited trait controlled by multiple genes. Overall, selecting lines with the shoot avoidance mechanism would be appropriate for As-contaminated regions since farmers in the major rice-growing regions of India and Bangladesh use rice straw to feed farm cattle (Azizur Rahman et al. 2008). Some of the lines from the population showing high tolerance and low accumulation of As in shoots could be used in breeding programs aiming to identify genes that could increase As tolerance with low accumulation (Additional file 1: Table S1) (Tuli et al. 2010).

3.5.2. Combined backcross breeding population

In this study, we exposed five founder lines WTR1(RP), Y 134 (DP6), Zhong 413 (DP7), OM 997 (DP9), BG 300 (DP10) and its 205 BC_1F_5 derived lines to a concentration of 5, 10, 15 ppm As stress for 18 days. When exposed to different concentration of As, a wide range of morpho-physiological changes occurred in the plants, leading to inhibited plant growth with a rapid decline in biomass and reduced chlorophyll content when compared to the control (control> 5 ppm As > 10 ppm As > 15 ppm As) (Figure 2.18 and 2.19) (Wu et al. 2011; Pandey et al. 2015). In our current experiment, the distribution of As showed similar patterns (root>shoot) regardless of treatment and in agreement with the previously reported genotype-dependent As accumulation (Abedin et al. 2002a). Results suggested that the *O. indica* donor parent DP7 (Zhong 413) accumulated less As in shoots than its other three donor and common recipient parents. The pattern among the traits under varying As treatment revealed that plant performance depended on As concentration in the tissue, with increased uptake in shoots reducing the plant‘s overall performance parent (Rahman et al. 2007; Bhattacharya et al. 2010; Wu et al. 2011). Arsenic toxicity traits observed in the rice population in our experimental conditions signify that As tolerance and accumulation in rice seedlings is a quantitatively inherited trait controlled by multiple genes.

3.5.3. As accumulation in the combined breeding population under field condition

Differences in As accumulation by rice under the same soil environmental conditions is necessary to reveal the genetic difference in rice (Kuramata et al. 2011). The distribution of As showed similar patterns (straw>husk>grain) regardless of the season grown and a significant difference was observed between the seasonal accumulation (Table 2.9 and Figure 2.21). In Dry season As accumulation was considerably high than the wet season possibly due to frequent irrigation in DS and result appears to be similar to that reported by (Ahmed et al. 2011). Field condition As accumulation also corresponding with As accumulation pattern of hydrophones (Khan et al. 2010; Bajpai and Upreti 2012; Liao et al. 2016). Lines derived from the cross between WTR1 × Zhong 413 accumulate less As in the grain, husk and straw possibly plants appear to restrict As uptake and translocation. Interrogation lines derived from cross WTR1 × OM997 and WTR1 × BG 300 accumulate more As in the straw and on the other hand accumulate less As in the grain. Lines derived from the cross WTR1 × Y134 accumulate more As in the grain than the other clusters. Overall mean As accumulation in grain was less than <0.20 mg kg^{-1} in the wet season and < 0.26 mg kg^{-1} in the dry season possibly due to water scares in dry season plants trend to uptake more available water along with As and other nutrients. However, some lines in the mapping population exceed >0.3 mg kg^{-1} As in the grain. Use of advanced breeding lines in mapping experiment is advantageous since lines are stabilised for yield-related traits few more rounds of testing is enough for varietal nomination (Ali et al. 2018b; Jernigan et al. 2018).

3.6. Genomic regions associated with As toxicity stress in the population

3.6.1. Single backcross breeding population

When using a 6K SNP-array for genotyping, the average number of polymorphic markers derived across diverse germplasm was quite high (Thomson et al. 2017). However, only 1068 SNP sites were identified in this study, with 704 polymorphic sites remaining after removing the monomorphic sites. As the population used in the current study was advanced using the backcross breeding approach, only a single fragment or a small number of genomic introgression fragments from a donor parent were retrieved in the population, leading to the lesser number of polymorphic sites present in the near-isogenic lines (Li and Ali 2017; Ali et al. 2018a). QTL mapping was

performed using 704 high-density SNP markers, and a total of 9 tightly linked QTLs associated with As-related traits were mapped using marker-trait association. The putative QTL regions detected for As-related traits in our study were compared with previous reports based on the physical positions of the associated markers in the Nipponbare genome (International Rice Genome Sequencing Project, http://rgp.dna.affrc.go.jp/IRGSP/). Associated SNPs mapped on chromosomes 2, 6, and 8 were consistent with previous studies for As content QTLs in rice (Dasgupta et al. 2004; Zhang et al. 2008; Jung et al. 2013). On the other hand, the QTLs mapped on chromosomes 1, 5, and 9 were reported for the first time in the As stress condition. Detection of multiple QTLs for traits in stable populations illustrates the complexity of As-induced traits (Pascual et al. 2016).

3.6.2. Combined backcross breeding population

In this study, it was demonstrate that the use of single and multiple parents in the development of population through single backcross and advancing using positive selection is advantageous in genetic mapping as it harnesses more allelic diversity, thereby enabling the detection of more precision QTLs as compared to using conventional bi-parental populations (Zhu et al. 2015; Ali et al. 2017). Sixteen QTLs were identified for As related traits and two QTLs for grain As content was mapped using the association analysis. QTLs mapped on chromosomes 2, 5, 6, and 8 were consistent with previous studies for As content QTLs in rice (Dasgupta et al. 2004; Zhang et al. 2008; Jung et al. 2013; Liu et al. 2019). Alternatively, the QTLs mapped on chromosomes 1, 3, 5, 11 and 12 were reported for the first time in the As stress. For investigating straighthead physiological disorder in rice, As was the main chemical used to induce phenotypes like sterile florets and distorted spikelets in rice (Yan et al. 2005; Rahman et al. 2008, 2012; Agrama and Yan 2009; Pan et al. 2012). However, the mechanism behind As induced sterility in rice spikelets was poorly understood. In our current mapping experiment, QTL-*qGrAsC*6 mapped on chromosome 6 for grain As content and QTL *qShAsC*11.1 mapped on chromosome 11 for shoot As content was colocalised with the previously identified QTL for the straighthead disorder in (Zhe733×R312) stable population of rice (Pan et al. 2012).

3.7. Candidate genes associated with Arsenic related traits

Plants have evolved versatile detoxification systems to counter the wide variety of phytotoxic compounds present in the natural environment (Coleman et al. 1997;

Shamsi et al. 2008). In a single breeding population, the region 39.28-39.42 Mb on chromosome 1, QTL *qRChlo1* was located. Of the four candidate genes, the most plausible one was *LOC_Os01g67580* (*OsMRP2*). In combined breeding population, the region 25.64-26.32Mb on chromosome 11, QTL *qShAsC11.2 was* located. Of the four candidate genes, the possible one was *LOC_Os01g67580* (*OsMRP3*) which encodes for a similar multidrug resistance-associated protein (MRP, ABCC1), a subfamily of ABC transporters. Members of the MRP ATP-binding cassette transporters were initially proposed to be primarily involved in the vacuolar sequestration of potentially toxic metabolites. MRPs exist in all organisms and are known to play critical roles in the efflux of xenobiotic compounds (Paumi et al. 2009). In one study involving *Arabidopsis thaliana* double knockout mutants of ABCC- type transporters (AtABCC1 and AtABCC2), the mutant plants exhibited a very low residual $As^{(III)}$–PC2 transport activity (Song et al. 2010). In another study, involving the same double knockout mutant lines (AtABCC1 and AtABCC2), tolerance of mercury and cadmium toxicity was conferred by sequestering in the large vacuole (Park et al. 2012). Another study, involving (AtABCC3) transporter mutant in *A. thaliana,* also confirmed the increasing sensitivity to cadmium toxicity in the seedlings (Brunetti et al. 2015). The gene *OsMRP* an ABCC1 transporter identified in the QTL region of *qRChlo1 and qShAsC11.2* may be involved in As transport and partitioning processes such as vacuolar sequestration of toxic metabolites by the action of ATP-driven efflux pumps (Klein et al. 2006; Jeandroz and Lamotte 2017).

QTLs associated with shoot As content included genes of complex physiological and molecular networks, ranging from metal transporters, sequestration and partitioning of As, and synthesis of detoxification proteins and cheaters that promote shoot-to-root mobility of As (Tripathi et al. 2013; Pandey et al. 2015). In population one, the region of 10.88-14.64 Mb on chromosome 5, QTL *qAsS*5.1 was located. Of the three candidate genes, the most likely one was *LOC_Os05g27100* annotated as heavy metal-associated domain-containing protein (HMA). In population two, the region of 19.37-19.39 Mb on chromosome 8, QTL *qShAsC*8 was located. Of the two candidate genes, the most likely one was LOC_Os08g31340 annotated as heavy metal-associated domain-containing protein (a Copper chaperone). HMAs are metal chaperones that contain a metal binding domain for the safe transport of metallic ions inside and outside of the cell, primarily involved in detoxification mechanism (De Abreu-Neto et al. 2013). In rice, gene *OsHMA9* is a metal efflux HMA protein known

to play a vital role in metal homeostasis of copper, zinc, lead, and cadmium (Lin and Liao 1979; Lee et al. 2007). In the case of As toxicity in rice, the role of HMA is yet to be defined.

In the region of 0.40-2.02 Mb on chromosome 6, QTL *qAsS*2 was located. Of the six candidate genes, the most likely one was *LOC_Os06g03540* (*OsOPT2*) annotated as oligopeptide transporter family (OPT), which plays a significant and diverse role in plant growth and development. It is widely accepted that OPTs are proton-coupled symporters that translocate their substrates in the cytosolic direction with the possible function of long-distance metal distribution, nitrogen mobilisation, heavy metal sequestration, and glutathione transport (Lubkowitz 2011). Studies suggest that 16 OPT gene motifs are distributed in the rice genome, and *OsOPT2* showed transcriptional accumulation in all tissues, with varying amounts under normal conditions and significantly down-regulated under salinity and cold stress and up-regulated under drought stress, suggesting that *OsOPT2* may be involved in diverse plant regulation functions (Liu et al. 2012). However, *OsOPT2* was not functionally characterised in rice. In population two, region 18.68-20.04 Mb on chromosome 11, two QTLs for different traits were detected (*qRph*11 and *qShAsC*11) and encodes three genes of the same function (*LOC_Os11g04530, LOC_Os11g04540* and *LOC_Os11g04550*) annotated as sulfotransferase domain-containing protein. Sulfotransferase plays a significant role in Arsenate reduction and regulation of accumulation in rice (Sánchez-Bermejo et al. 2014). Arsenate reductase activity was confirmed in an arsenate-hyper accumulating fern. The reaction mechanism was very similar to the previously reported activity of ACR from yeast, using GSH as the electron donor (Duan 2005; Ellis et al. 2006).

In the region of 4.34-7.22 Mb on chromosome 2, QTL *qAsS*2 was located. Of the 13 candidate genes, the most likely one was *LOC_Os02g09150*, which encodes for inorganic H+ vacuolar pyrophosphates (V-PPase). Plant V-PPase is a specific class of multi-subunit pumps that play an essential role in the productivity of higher plants, generation of proton gradients across tonoplast endomembrane, and their ability to buffer changes in the concentrations of essential and toxic ions (Bak et al. 2013). A study involving *A. thaliana* double knockout mutant lines of V-PPase (vha-a2 and vha-a3) showed reduced zinc tolerance with no response to salt stress (Krebs et al. 2010). Another study involving the same double knockout mutant lines (vha-a2 and vha-a3) displayed functional regulation of intracellular ion homeostasis by ion

compartmentalisation of calcium ion (Tang et al. 2012). However, their functional relation and relative contributions to ion storage and detoxification are unclear in rice. In most cases, including root-related metal concentration traits, excess mineral attached to the root surface may interfere with the measurement. This measurement is not considered as a good-quality trait for understanding mineral uptake and metabolism in rice (Zhang et al. 2008). Two QTLs identified for root As content, *qAsR*8.1 (5.84-6.28 Mb) and *qAsR*8.2 (7.33-8.38 Mb), were located 1 Mb apart from each other on chromosome 8, and co-localized with a QTL previously reported for As content in brown rice (Zhang et al. 2008). Within the region was harboured the well-studied rice gene (*OsZIP4*) responsible for zinc transport and distribution in rice (Ishimaru et al. 2005, 2007). Moreover, genes present in the root QTL intervals for As content did not show any response in the gene expression analysis.

4. CONCLUSION

Rice is a significant source of dietary intake of arsenic for the populations that consume rice as a staple food and reducing the levels of carcinogenic arsenic in rice is a primary public health goal. In the present study, the identification of arsenic-tolerant and excluder rice genotypes provides the basis for the development of permanent breeding populations. Results obtained from this study provide novel insight into the genetic basis of As tolerance and uptake in the rice. Also, this study discussed the utility of tunable genotyping-by-sequencing (tGBS®) technology in eleven-backcross breeding populations.

In conclusion, our study revealed the availability of new sources for arsenic tolerance and low accumulating., which will be undoubtedly helpful for the feature rice-breeding program. Genotyping with developed permanent breeding lines substantiated the impacts of novel breeding strategy revealed: (a) the donor introgression patterns in populations were characteristic with variable introgression frequency in different genomic regions, attributed mainly to stringent selection under abiotic stress (b) considerably lower heterozygosity observed in developed lines signifying the stability of developed populations. Through marker-trait association analysis, a total of twenty-five QTLs were identified on the different chromosomes in rice. Demonstrating arsenic tolerance and uptake is complex quantitatively inherited trait controlled by multiple genetic factors. Comparing our findings with the previously reported QTLs for As toxicity stress in rice, we identified some novel and co-localized QTLs associated with As stress. Two major QTLs *qGrAsC*6 and *qShAsC*8 mapped for arsenic content in grain and shoot respectively were consistent with previous studies, and pyramiding lines containing this QTLs may be useful in the breeding for arsenic tolerance. Also, the mapped QTLs harbour gene models of known function associated with stress responses, metal homeostasis, and transporter activity in rice. These results advance the understanding of genetic factors associated with arsenic tolerance and uptake in rice. Overall, our findings will assist rice breeders with initial marker information to develop suitable varieties for arsenic-contaminated ecosystems.

5. MATERIALS AND METHODS

5.1. Plant materials

5.1.1. Breeding panel

Fifty-three genotypes consisting of 7 high yielding GSR lines and 46 different rice cultivars, from the core collection of IRRI-GSR breeding program (Table 5.1), were initially screened to evaluate As tolerance and accumulation. Based on contrasting tolerance response patterns, genotypes were selected to develop mapping populations for dissecting genomic regions associated with As-tolerant traits. These 46 rice cultivars originated from major rice-growing areas of China, India, Philippines, Vietnam, Bangladesh, Sri Lanka, Pakistan, United States of America, Brazil, and Iran, where As contamination has been reported in paddy soil (Bastías and Beldarrain 2016; Bakhat et al. 2017). Besides, these diverse cultivars include the parents of several permanent mapping populations developed in IRRI for yield improvement under unfavourable environments.

Table 5.1: Description of plant cultivars used for Arsenic tolerance screening

Cultivar name	Subspecies	Origin country	Cultivar name	Subspecies	Origin country
GSR 8	*Indica*	Philippines	Jhona 349	*Intermediate type*	India
GSR IR1-12-Y4-D1-Y2	*Indica*	Philippines	Binam	*Japonica*	Iran
GSR 12	*Indica*	Philippines	Basmati 385	*Intermediate type*	Pakistan
GSR IR1-17-Y16-Y3-S1	*Indica*	Philippines	M401	*Temperate japonica*	United States
GSR IR1-5-Y4-S1-Y1	*Indica*	Philippines	X21	*Indica*	Vietnam
GSR IR1-5-Y7-Y2-SU1	*Indica*	Philippines	IRAT 109	*Tropical japonica*	Brazil
GSR IR1-4-S5-L1-L1	*Indica*	Philippines	Gayabyeo	*Intermediate type*	Korea
OM1706	*Indica*	Vietnam	Shwe-Thwe-yin-Hye	*Intermediate type*	Myanmar
IR64	*Indica*	Philippines	NAN29-2	*Intermediate type*	India
OM1723	*Indica*	Vietnam	FR13-A	*Indica*	India
IRAT352	*Indica*	Philippines	Jiangxi-Si-Miao	*Indica*	China
Phalguna	*Indica*	India	TKM9	*Indica*	India
Zhong413	*Indica*	China	PSBRc82	*Indica*	Philippines
Teqing	*Indica*	China	NPT	*Indica*	Philippines
BR24	*Indica*	Bangladesh	Huang-Hua-Zhan (HHZ	*Indica*	China
PSBRc66	*Indica*	Philippines	Hau 565	*Indica*	China
CDR22	*Indica*	China	Bg304	*Indica*	Sri Lanka
WTR 1	*Indica*	Philippines	Kalimekri 77/5 (6527)	*Indica*	China
Feng-Ai-Zhan	*Indica*	China	BR11	*Indica*	Bangladesh
Qi-Si-Ying-Zhan	*Indica*	China	IR66	*Indica*	Philippines
Xing-Ying-Zhan	*Indica*	China	Thadokkham 1 (TDK1)	*Indica*	India
Te-Huang-Zhan	*Indica*	China	Feng Fu Zhan (FFZ)	*Indica*	China
Y-134	*Indica*	China	TME80518	*Indica*	China
Haoannong	*Japonica*	China	Zhongzu 14 (ZZ14)	*Japonica*	China
Cheng-Hui 448	*Indica*	China	SACG-4	*Indica*	China
Bg300	*Intermediate type*	Sri Lanka			
OM997	*Aus*	Vietnam			
Basmati	*Indica*	India			
Jhona 349	*Intermediate type*	India			

5.2. Mapping population development

5.2.1. Single backcross breeding population

A BC_1F_2 population was established from a cross between an elite Xian (*Indica*) variety, Weed Tolerant Rice1 (WTR1, as the recipient) and a Chinese Geng (*Japonica*), Hao-An-Nong (HAN as the donor). WTR1 is a high yielding variety from south China with wide adaptability across subtropical and tropical areas of Asia. WTR1 was crossed with HAN in the 2010 dry season (DS) and the F_1 plants were backcrossed to WTR1 once in the 2010 wet season (WS). In the 2011 DS, seeds from more than 25 segregating BC_1F_1 plants were bulk harvested without selection to form a single BC_1F_2 population. Phenotypic selection schemes of the BC_1F_2 bulk population to develop lines as illustrated in (Figure 5.1). With improved nutrient use efficiency at the IRRI field station at 14°13′ N and 121°15′ E, at an elevation of 21 m above mean sea level in the Philippines during the three wet seasons of 2011-2013 and two dry seasons of 2012 and 2013, respectively. The first round of single-plant selection for higher grain yield was practised on the BC_1F_2 bulk population grown under irrigated conditions (IG), low-input (LI) and rainfed (RF) situations in 2011 WS, resulting in 46 selected BC1F2 plants. Then, in the following seasons, four rounds of line-based selection were carried out from progeny testing under six different nutrient input conditions, including the NPK (Nitrogen, Phosphorus, Phosphorus), 75N, -N, -P, -NP, and –NPK conditions. The NPK condition was the condition with total applied NPK fertilisers of 160-50-50 kg/ha^{-1} in the DS and 90-30-30 kg/ha^{-1} in the WS. The 75N, -N, -P, -NP, and -NPK conditions indicated that 75% of nitrogen, the absence of N, the absence of P, the absence of both N and P, and finally absence of NPK were used, respectively, compared with the standard NPK condition. In each season, seeds of each selected line and parents were sown on the seedling nursery, and 20 to 24-day old seedlings of each line were transplanted into a two-row plot at a spacing of 20 cm × 15 cm with one seedling per hill. Plant selection was practised by selecting the best 1-3 plants from the best yield performing lines.

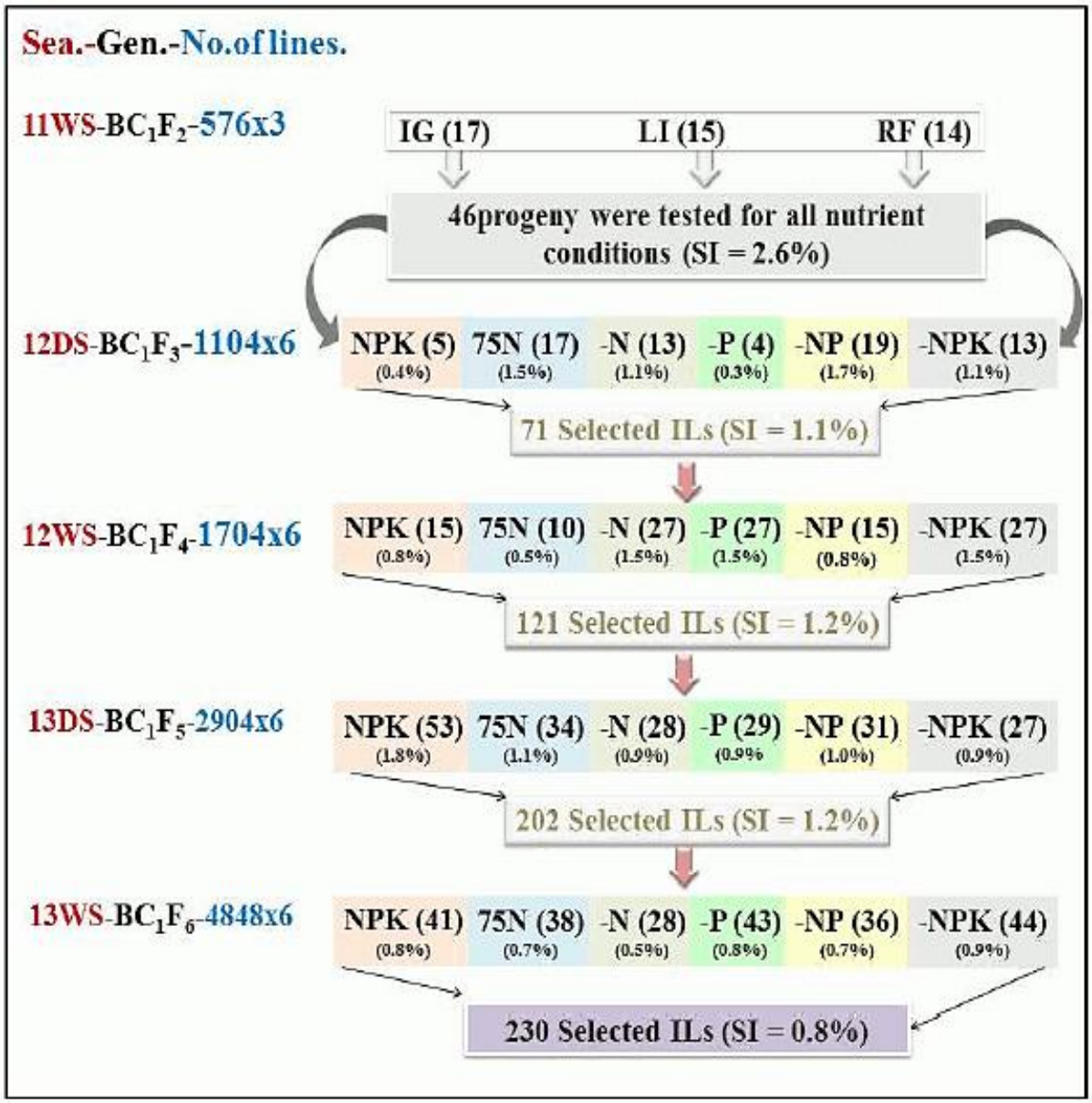


Figure 5.1. Phenotypic selection schemes of the BC_1F_2 bulk populations to develop 230 permanent breeding lines.

Sea- rice growing season, Gen-generation, WS- wet season, DS- dry season, BC-back cross, F-filial generation, IG-irrigated condition, LI- low input condition, RF-rainfed condition, SI-selection intensity, ILs-introgression lines, N-Nitrogen, P-Phosphorus, K- Potassium, 75N- 75 percent Nitrogen application and negative sign indicates particular nutrient or in combination was not supplied. The experimental field was maintained for six consecutive years without fertiliser application.

5.2.2. Interconnected breeding population

Eleven donors (elite cultivars selected from different rice agro-ecologies) were crossed with Weed Tolerant Rice-1 (WTR-1), and their F_1s were backcrossed once with WTR-1 (Table 5.2). Subsequently, the F_1BC_1s were self-pollinated, their seeds were bulked to create 11 BC_1F_2 populations that were screened for three rounds beginning in the 2011 wet season under various biotic and abiotic stresses such as drought, low input, salinity, submergence, tungro, and standard irrigated conditions, among others. In all environments, surviving plants that exhibited superior performance over the checks and the WTR-1 recipient parent (RP) were selected using a previously published approach as illustrated briefly in (Figure 5.2). This novel early-backcross breeding technique has been proven successful in exploiting favourable genes hidden in diverse germplasm to develop lines that are tolerant to multiple stresses (Ali et al. 2006, 2012a, b, 2013). Based on the parent performance of the 11 donor parents under As stress. Only four out of eleven developed populations of BC_1F_5 backcross lines were considered for mapping experiment. Lines derived from the cross between common recipient parent Weed Tolerant Rice-1 (*IndX*) with four elite donor cultivars Y 134 (*Ind1B*) 34 lines, Zhong 413 (*IndX*) 56 lines, BG 300 (*IndX*) 60 lines, OM 997 (*Ind1B)* 55 lines were combined for phenotyping under As toxicity stress.

Table 5.2: Description of plant cultivars used in population development

Subpopulation*	Donor parent (DP) and source country
1	Haoannong (DP1)-China
2	Cheng-Hui 448 (DP4)-China
3	Feng-Ai-Zan (DP5)-China
4	Y 134 (DP6)- China
5	Zhong 413 (DP7)-China
6	Khazar (DP8)-Iran
7	BG 300 (DP9)-Srilanka
8	OM 997 (DP10)-Vietnam
9	Basmati-385 (DP12)-Pakistan
10	M 401 (DP17)-USA
11	X 21 (DP19)-Vietnam
Total	**11 Donor parents**

*All sub-populations have common a recipient parent, Weed Tolerant Rice-1(WTR-1)

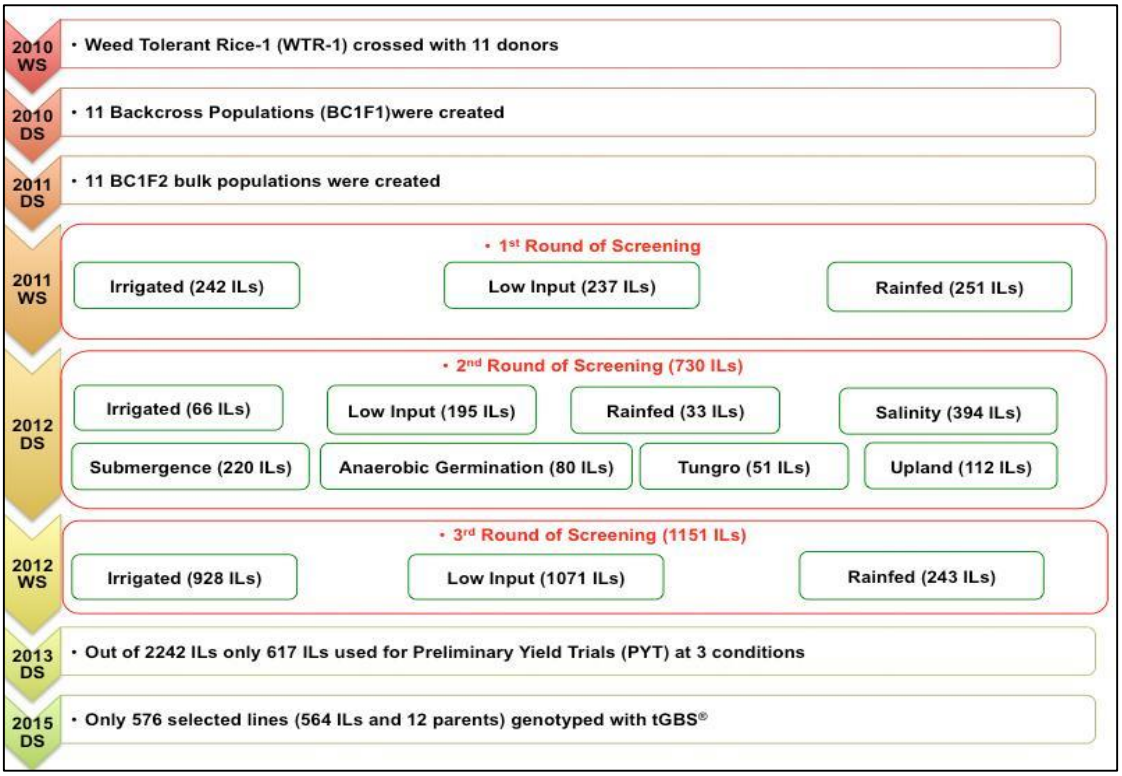


Figure 5.2. Phenotypic selection schemes of the BC_1F_2 bulk populations to develop 576 permanent breeding lines.

Ds-Dry season, Ws-Wet season and ILs- interrogation lines, WTR1- Weed Tolerant Rice-1, BC-back cross, F-filial generation, tGBS- tunable genotyping by sequence.

5.3. Phenotyping

5.3.1. Germination assay

Germination screening assay was carried out in the plant growth chamber facility at IRRI for evaluating the germination tolerance of genotypes under varying As treatments. Five levels of inorganic As treatment, 0 ppm (Control), 5 ppm (low), 10 ppm (medium), 15 ppm (high) and 20 ppm (very high) was supplied in the form of Sodium Arsenate ($Na_2HAsO_4.7H_2O$, Sigma Aldrich, Singapore) to evaluate the germination ability of seeds under As toxic condition. All the purified seeds of the respective cultivars where surface sterilised with 1% Sodium hypochlorite (NaOCl) for one minute followed by rinsing with deionised water for several times. Germination was evaluated on a moist filter paper (Whatman No.1) bed dampen with 10 ml of respective As treatment placed on each of the 15 cm diameter Petri dishes. Fifty sterilised seeds per genotype were laid on each petri dish and incubated at the room temperature (28° - 32 °C). Each treatment was replicated three times and seeds were allowed to germinate for ten days. During this period, the Petri dishes were moistened

with respective solutions of As when needed. After ten days of incubation, the number of seeds germinated was recorded by the emergence of the radical and coleoptile at the respective treatments and controls. Based on the tolerance percentage under different levels of As treatments, genotypes were ranked into four different categories, highly tolerant (>80% germination), moderately tolerant (>50% germination), moderately susceptible (<50% germination) and highly susceptible (<20% germination). The tolerance percentage of cultivars was calculated with the following equation.

$$\text{Tolerance percentage} = \frac{\text{Germination \% in each treatment}}{\text{Germination \% in the control}} \times 100$$

5.3.2. Hydroponics culture and screening

5.3.2.1. Breeding panel

Arsenic toxicity screening experiments were conducted in a hydroponic system in the controlled phytotron glasshouse of IRRI during 2014 wet season. Optimum rice growing conditions were maintained throughout the experiment, i.e., 29/21°C (day/night), 70% relative humidity and natural light. Seeds of the 53 lines were oven dried for five days at 60 °C to break any residual seed dormancy and incubated at 30°C for 48 h for germination. One seedling per line was transferred per hole with 1 cm diameter on a Styrofoam seedling float with the size of 28×32×1.25 cm having 100 holes (10×10) with nylon net bottom fixed in a dark plastic tray containing 8 L of full strength Yoshida nutrient solution (Gregorio et al. 1997; Wu et al. 2016). Throughout the experiment, the pH range was maintained in a range of 5.1 to 5.4 in the nutrient solution. At the 7th day, 5 ppm Arsenic, 10 ppm Arsenic and 15 ppm Arsenic was supplied in the form of Sodium (Meta) Arsenite ($AsNaO_2$, Sigma-Aldrich, MO, USA). A control treatment was maintained throughout the experiment. Plants were grown in Arsenic toxic condition for 18 days, and nutrient solutions were renewed once every seven days. The experiment was laid out as a complete randomised design with three independent replicates, and five repeats per lines in each replicate, leading to thirty-two hydroponic tanks, each accommodating up to 100 seedlings.

5.3.2.2. Backcross breeding population

For the hydroponic screening experiments, 194 lines from the developed population were used, and the experiment was carried out in a controlled phytotron glasshouse

of IRRI during 2015 wet season. Optimum rice growing conditions were maintained throughout the experiment, i.e., 29/21^{0}C (day/night), 70% relative humidity and natural light. Seeds of the 194 lines were oven dried for five days at 60 °C to break any residual seed dormancy and incubated at 30 °C for 48 h for germination. One seedling per line was transferred per hole with 1 cm diameter on a Styrofoam seedling float with the size of 28×32×1.25 cm having 100 holes (10×10) with nylon net bottom fixed in a dark plastic tray containing 8 L of full strength Yoshida nutrient solution (Gregorio et al. 1997; Wu et al. 2016). Throughout the experiment, the pH range was maintained in a range between 5.1 to 5.4 in the nutrient solution. At the 7th day, 10 ppm As was supplied in the form of Sodium-(Meta)-Arsenite ($AsNaO_2$, Sigma-Aldrich, MO, USA). A control group was maintained throughout the experiment. Plants were grown in As toxic condition for 18 days, and nutrient solutions were renewed once every seven days. The experiment was laid out as a complete randomised design with three independent replicates, and five repeats per lines in each replicate, leading to fifty-nine hydroponic tanks, each accommodating up to 100 seedlings.

5.3.2.3. Interconnected backcross breeding population

Hydroponic screening experiment for As tolerance and uptake at the seedling stage were carried out in a controlled phytotron glasshouse of IRRI during 2016 DS. The mapping population used in this experiment consisted of 210 lines including five parents. Optimum rice growing conditions were maintained throughout the experiment. i.e., 29/21 ^{0}C (day/night), 70% relative humidity and natural light. Seeds of the 205 lines and five parents were oven dried for five days at 60 °C to break any residual seed dormancy and incubated at 30 °C for 48 h for germination. One seedling per entry was transferred per hole with 1 cm diameter on a Styrofoam seedling float with the size of 28×32×1.25 cm having 100 holes (10×10) with nylon net bottom fixed in a dark plastic tray containing 8 L of full strength Yoshida nutrient solution (Gregorio et al. 1997; Wu et al. 2016). Throughout the experiment, the pH range was maintained in a range between 5.1 to 5.4 in the nutrient solution. At the 7th day, 5 ppm As, 10 ppm As, 15 ppm As was supplied in the form of Sodium-(Meta)-Arsenite ($AsNaO_2$, Sigma-Aldrich, MO, USA). A control treatment was maintained throughout the experiment for each treatment. Plants were grown in Arsenic toxic condition for a total of 21 days, and nutrient solutions were renewed once every seven days. The experiment was laid out as a complete randomised design with three independent replicates, and five

repeats per lines in each replicate, leading to a total of 126 hydroponic tanks, each accommodating up to 100 seedlings.

5.3.3. Chlorophyll concentration and seedling growth parameters

To compare the differences among the lines the relative chlorophyll concentrations were measured non-destructively from the base, middle, and tip of the uppermost leaves of each individual plant, and the average values were expressed as SPAD units (SPAD-502 chlorophyll meter Minolta Camera Co., Ltd., Japan) as the indicator of leaf senescence caused by toxic As treatment. Plants' responses to the As treatment were evident after 18 days PAT. Changes in the shoot and root length response to the Arsenic treatment were measured for each entry at 18 days PAT. Shoot length was measured from the base of the plant to the tip of the longest leaf, while the root length was measured from the base of the plant to the root tip. Three plants per entry per replicate were rinsed with deionised water to remove any excess nutrients sticking to the surface of the root and then oven dried for three days at 70^0C to remove moisture. Before sample processing for As analysis, dry biomass was recorded.

5.3.4. Field screening for As content

Grain As content screening experiment was conducted at the field station of IRRI. For breeding panel in dry/wet seasons of the year 2013, backcross breeding population As content in grain was not determined, since the typhoon Rammasun destroyed the collected samples in July 2014 and for interconnected breeding, population experiment was carried out in dry and wet seasons of the year 2015. IRRI is situated at latitude 14°13'N and longitude 121°15'E with paddy soil type is a Maahas clay loam, isohyperthermic mixed type tropical with average total As content in soil ranged from 2.6 to 7.2 mg kg^{-1} of dry soil and irrigation water As concentration ranged from 0.021 to 0.047 mg l^{-1} (data not shown). Most natural soils contain low levels of As, Background concentrations in agriculture soil range from 1 to 40 mg kg^{-1}, with a mean value of 5 mg kg^{-1} (Abernathy et al. 2001; Wong et al. 2012). As concentration, existing in the IRRI soil was suitable for studying As accumulation in plants under average condition. For each season 21-day, old seedlings were transplanted to the field with experiment design laid out in randomised complete block design into two-row plots with 12 plants per row at a spacing of 20 cm × 20 cm. Standard agronomic practices with optimum fertiliser, irrigation, and plant protection measures were adept at

ensuring good crop growth cycle for grain development. Matured filled panicles along with straw for each cultivar were bulk –harvested from the eight central plants from each replication. For Arsenic content analysis, the harvested seeds were de-husked, separated into brown unpolished rice and husk. Separated grain samples were stored separately for three months before analysis.

5.3.5. Analytical procedure for determining As concentration

Arsenic content in roots and shoot sample from the hydroponic screening experiment, grain, husk and straw samples from the field experiment were analysed for total As content. Dried samples were thoroughly homogenised by an ultra-centrifugal mill (ZM100, Retsch, Haan, Germany) modified with a tungsten blade to avoid any cross-contamination. Grounded 0.2 to 0.5 grams of sample were added to a closed vessel digester, 5 mL of high-purity 69 % concentrated Nitric acid (HNO_3) followed by 2 mL Hydrogen peroxide (H_2O_2), 1 mL of deionised water was added and pre-digested overnight in the fume hood (Amaral et al. 2013). The following day the samples were digested using a heating block at 150- 155 ^{0}C for three 3 hours under the fume hood. The digested tissue was diluted to a final volume of 25 mL using deionised water and total As was determined using Graphite Furnace Atomic Absorption Spectrophotometry (GFAAS-7000F, SHIMADZU corporation, Kyoto, Japan) (Environmental Protection Agency USEPA 1998). A deuterium lamp background correction with high-performance As hollow cathode lamps (193.7 nm) was used to quantify the As content in the standards and samples. An aliquot of 20μL of the digested sample and 10 μL Palladium (100 ppm) as matric modifier were injected into the pyrol-coated graphite tube with the aide of an autosampler unit. Seven-stage furnace cycle program was adopted with an atomization temperature of 2200 ^{0}C at the 6th stage were used as a standard parameter to determine As content in the sample and expressed in milligram per kilogram (mg kg^{-1}).

5.3.6. Statistical analysis

Two-way ANOVA was carried out to observe the effects of lines, treatment and lines-treatment interaction on different As-related traits. For breeding panel screening experiment, the principal component analysis was also carried out to observe the pattern of variation among the genotypes, the relationship among individuals and their characteristics. As content from the field experiment, Student t-test was carried out to observe the variation in As accumulation between dry seasons. For mapping analysis,

the relative phenotypic value of morphological traits was determined for each genotype, i.e., the phenotypic value obtained from the As treatment divided by control value (Dasgupta et al. 2004; Shrestha et al. 2018). Pair-wise Pearson‘s correlation analysis was carried out among the measured traits, in which *P*- value was two-tailed with two significant levels were using *P*=0.05 and *P*=0.01. All the statistical analysis were conducted using PBTools package of R (R Core Team 2015).

5.4. Genotyping

5.4.1. Genome-wide 6K SNP array

Genomic DNA was extracted from the parents and population using the DNeasy Plant Mini Kit following the manufacturer's protocol (QIAGEN, Germantown, MD, USA). High-throughput SNP genotyping was carried out using a custom design Illumina Infinium rice 6K-SNP array containing 4606 SNPs covering all 12 rice chromosomes at the Genotyping Services Laboratory facility at IRRI (Thomson 2014; Thomson et al. 2017). A rice 6K-SNP chip was scanned using an Illumina bead array reader, and automatic allele calling was achieved using the Illumina Genome Studio data analysis software (V2010.1. Illumina Inc)(Thomson et al. 2017). Scoring of alleles in the BRILs at each SNP locus was carried out by comparing parental alleles at the particular SNP locus as co-dominant markers.

5.4.2. Tunable genotyping by sequencing

All 576 lines from the interconnected breeding population were sequenced using 10 Ion Proton runs. The rice reference genome (Osativa_204_v7.0.fa) was downloaded from the Phytozome website (https://phytozome.jgi.doe.gov/pz/portal.html). The typical workflow of Tunable genotyping-by-sequencing (tGBS®) analysis with a reference genome is illustrated in (Figure 5.3). All the DNA samples were digested using two restriction enzymes (*Nsp*I and *Bfu*CI). Ligation followed with a single-stranded barcode oligonucleotide in one site and an oligonucleotide complementary to amplification primer in the other site (Islam et al. 2015). Trimming of raw sequence reads was conducted using Lucy software (Chaou et al. 1998; Li and Chou 2004). Nucleotides at each site were scanned for low quality and the bases with a PHRED quality value of <15, i.e., error rates of ≤3%, were removed (Ewing and Green 1998). Trimmed reads were aligned to the public reference genome using GSNAP (Wu and Nacu 2010; Islam et al. 2015). Mapped reads were used for SNP discovery (≤2 mismatches for every 36 bp and <5 bases for every 75 bp as unaligned tails).

Polymorphisms at each potential SNP site were carefully examined (supported by at least three reads) and putative homozygous and heterozygous SNPs were identified.

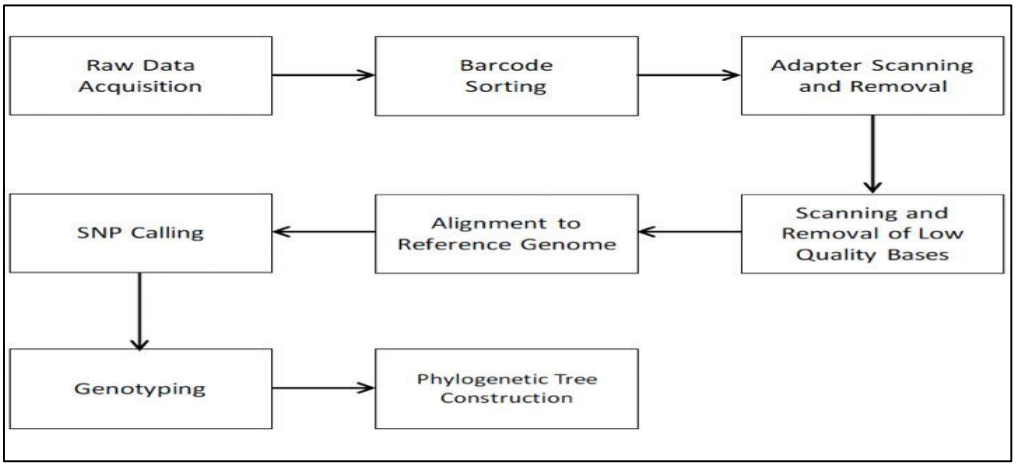


Figure 5.3. Workflow of tunable genotyping by sequence analysis with a reference genome (Osativa_204_v7.0.fa).

Homozygous SNPs were called following the criteria of having a PHRED base quality of 20 (≤1% error rate) and at least three reads supporting the common dominant allele. Heterozygous SNPs were called if at least two reads supported each of at least two different alleles, and each of the two read types separately comprised >20% of the reads aligning to that site, and also, if the sum of the reads supporting those two alleles comprised at least 90% of all reads covering the site. SNP calls were then filtered based on a missing data rate of ≤80% having an allele number of 2, the number of genotypes of ≥2, the minor allele frequency of ≥0.1, and heterozygosity range of 0-10% (Islam et al. 2015; Leiboff et al. 2015). Finally, the LMD50 (low missing data) SNP dataset was defined by filtering again based on a missing data rate of ≤50%. For each SNP, a major allele and a minor allele were defined based on the “reference (REF) allele: alternative (ALT) allele” ratio in each sub-population. If for a given SNP, the ratio was >1, the REF and ALT alleles were designated as the major and minor alleles, respectively. In contrast, if the ratio was <1, the REF and ALT alleles were designated as the minor and major alleles, respectively.

5.5. QTL mapping analysis

5.5.1. Backcross breeding population

QTL mapping for As-related traits from bi-parental was carried out using the software iciMapping v.4.0 (http://www.isbreeding.net/software/?type=detail&id=18). Trait-SNP associations were detected using the function single marker analysis (SMA) of the

iciMapping program under the bi-parental mapping function. The LOD threshold was set based on a permutation test (1000 permutations, P = 0.01) for each trait (Doerge and Churchill 1996). The confidence interval of each QTL was delimited by the linkage disequilibrium (LD) which was calculated by a genetics package in R software (Pang et al. 2017; Paradis et al. 2017). Significant QTL for each trait was named following the standard protocol (McCouch 1997). The additive effect was computed as the AA − (AA + BB)/2, where AA and BB is the mean phenotype value of genotype AA and BB, respectively. The phenotype variance explained (PVE) for each QTL was estimated by $1 - 10^{-2LOD/n}$ where n is the sample size, and LOD is the logarithm (base 10) of odds score obtained (Pang et al. 2017).

5.5.2. Interconnected breeding population

Association analysis was conducted in the combined four selected backcross breeding populations using a mixed linear model (MLM) by treating the founder effects of each locus as fixed effects. The population structure was considered by defining a $n \times 4$ matrix of founder allele inheritance indicator for locus k, and the kinship matrix was used to estimate the polygenic effect to reduce spurious association, Structure analysis was performed using program Structure (version 2.3.4) (Zhu et al., 2015; Wei and Xu, 2016). Let y be a $n \times 1$ vector for the phenotypic values of n individuals and Z_k is a $n \times 4$ matrix of the founder allelic indicators for the locus k. The j^{th} row of the matrix Z_k is a 1×4 vector of allelic indicators for the individual j. If the individual is a homozygote and both alleles from the first founder, the Z_{jk} is defined as $Z_{jk} = [2\ 0\ 0\ 0]$. If the individual is a heterozygote carrying the second and third alleles, the Z_{jk} is defined as $Z_{jk} = [0\ 1\ 1\ 0]$. The sum of all four elements in Z_{jk} equals to 2. The mixed model for testing the significance of the k^{th} marker is defined as

$$y = X\beta + Z_k\gamma_k + \xi + \varepsilon$$

Where Z_k the inheritance indicator for the marker is k, γ_k is a 4×1 vector for the four founder's allelic effects, ξ Is the polygenic effect and ε is the residual error. The threshold to declare a significant association was set at a probability level of 1.0×10^{-4} (Wissuwa et al. 2015).A Linkage disequilibrium (LD) block harbouring significant SNPs

were defined as a putative QTL region, and analysis was carried out using the computer program Haploview version 4.2 (Barrett et al. 2005).

5.6. Candidate gene analysis

For the identified QTLs governing As-related traits from the two populations, the gene models located in their confidence interval were searched from the MSU Rice Genome Annotation Database (Ouyang et al. 2007). In the identified QTL region loci showing non-synonymous polymorphism between the parents were considered for the candidate genes nomination (Pang et al. 2017). Polymorphisms between parental lines in QTL regions were retrieved from the Rice SNP-Seek Database (http://snp-seek.irri.org) (Alexandrov et al. 2015). For the backcross breeding population, additional genotype information of parents was obtained from an earlier tGBS® project using 10 Ion Proton runs, resulting in 794,297 polymorphic SNPs (https://doi.org/10.7910/DVN/RRXCR3) (Ali et al. 2018a). Raw sequence reads of parents obtained from the tGBS® projects were aligned to the public reference rice genome Osativa_204_v7.0.fa, downloaded from Phytozome website (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_%20Osativa). By using Genomic Short-read Nucleotide Alignment program (GSNAP) (http://research-pub.gene.com/gmap/), short reads are confidently mapped to the best location in the reference genome by allowing ≤2 mismatches every 36 base pairs (bp) and less than 5 base pairs for every 75 bp as tails were used as criteria. Aligned reads of the parents were compared with the reference genome sequence for the polymorphic site and compared among the parents to obtaining the SNP information (Wu and Nacu 2010; Ott et al. 2017). Then we selected gene models in the QTL interval showing non-synonymous polymorphism associated with metal homeostasis as our most likely candidate genes. Expression profiles of the most likely candidates in the identified QTL region was extracted from the previously published study, in which the rice cultivar Nipponbare (*Oryza sativa L. ssp japonica*) had been exposed to sodium Arsenite (As^{III}) stress for 24hr in the seedling stage (Yu et al. 2012). Data from that expression profile experiment was publically available (Xia et al. 2017; BIG Data Center Members 2018).

6. REFERENCES

Abedin J, Cresser MS, Meharg AA, et al (2002a) Arsenic accumulation and metabolism in rice (*Oryza sativa* L.). Environ Sci Technol 36:962–968

Abedin MJ, Cresser MS, Meharg AA, et al (2002b) Arsenic Accumulation and Metabolism in Rice (*Oryza sativa* L.). Environ Sci Technol 36:962–968. doi: 10.1021/es0101678

Abedin MJ, Feldmann J, Meharg AA (2002c) Uptake Kinetics of Arsenic Species in Rice Plants. PLANT Physiol. doi: 10.1104/pp.010733

Abedin MJ, Meharg AA (2002) Relative toxicity of arsenite and arsenate on germination and early seedling growth of rice (*Oryza sativa L*.). Plant Soil 243:57–66

Abernathy C, Chakraborti D, Edmonds JS, et al (2001) Environmental health criteria for arsenic and arsenic compounds. Environ. Heal. Criteria

Agrama HA, Yan WG (2009) Association mapping of straighthead disorder induced by arsenic in *Oryza sativa*. Plant Breed. doi: 10.1111/j.1439-0523.2009.01631.x

Ahmed ZU, Panaullah GM, Gauch H, et al (2011) Genotype and environment effects on rice (*Oryza sativa L.*) grain arsenic concentration in Bangladesh. Plant Soil 338:367–382. doi: 10.1007/s11104-010-0551-7

Akpinar BA, Lucas S, Budak H (2017) A large-scale chromosome-specific SNP discovery guideline. Funct Integr Genomics 17:97–105. doi: 10.1007/s10142-016-0536-6

Alexandrov N, Tai S, Wang W, et al (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res 43:D1023–D1027

Ali a. J, Xu JL, Ismail a. M, et al (2006) Hidden diversity for abiotic and biotic stress tolerances in the primary gene pool of rice revealed by a large backcross breeding program. F Crop Res 97:66–76. doi: 10.1016/j.fcr.2005.08.016

Ali J, Aslam UM, Tariq R, et al (2018a) Exploiting the genomic diversity of rice (*Oryza sativa* L.): SNP-typing in 11 early-backcross introgression-breeding populations. Front Plant Sci 9:1–10. doi: 10.3389/fpls.2018.00849

Ali J, Aslam UM, Tariq R, et al (2018b) Exploiting the Genomic Diversity of Rice (*Oryza sativa* L.): SNP-Typing in 11 Early-Backcross Introgression-Breeding Populations. Front Plant Sci 9:1–10. doi: 10.3389/fpls.2018.00849

Ali J, Jewel ZA (2018) Molecular Genetics and Breeding for Nutrient Use Efficiency in Rice. 1–27. doi: 10.3390/ijms19061762

Ali J, Xu J-L, Gao Y-M, et al (2017) Harnessing the hidden genetic diversity for improving multiple abiotic stress tolerance in rice (*Oryza sativa* L.). PLoS One 12:e0172515. doi: 10.1371/journal.pone.0172515

Ali J, Xu JL, Gao YM, et al (2013) Breeding for yield potential and enhanced productivity across different rice ecologies through green super rice (GSR) breeding strategy. Int Dialogue Percept Prospect Des Rice 60–68

Ali J, Xu JL, Gao YM, et al (2012a) Innovative Green Super Rice (GSR) molecular breeding strategy: achievements and advances. Philipp J Crop Sci

Ali J, Xu JL, Gao YM, et al (2012b) Green super rice (GSR) technology: an innovative breeding strategy-achievements & advances. In: Proceedings of ‘‘The 12th SABRAO congress-plant breeding towards. pp 16–17

Amaral CDB, Nóbrega JA, Nogueira ARA (2013) Sample preparation for arsenic speciation in terrestrial plants - A review. Talanta 115:291–299

Amini M, Abbaspour KC, Berg M, et al (2008) Statistical modeling of global geogenic arsenic contamination in groundwater. Environ Sci Technol. doi: 10.1021/es702859e

Anawar HM, Akai J, Komaki K, et al (2003) Geochemical occurrence of arsenic in groundwater of Bangladesh: Sources and mobilization processes. J Geochemical Explor. doi: 10.1016/S0375-6742(02)00273-X

Arai-Kichise Y (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. Plant Cell Physiol 52:274–282

Ashikari M, Ma JF (2015) Exploring the power of plants to overcome environmental stresses. Rice. doi: 10.1186/s12284-014-0037-y

Azizur Rahman M, Hasegawa H, Mahfuzur Rahman M, et al (2007) Effect of arsenic on photosynthesis, growth and yield of five widely cultivated rice (*Oryza sativa* L.) varieties in Bangladesh. Chemosphere. doi: 10.1016/j.chemosphere.2006.11.061

Azizur Rahman M, Hasegawa H, Mahfuzur Rahman M, et al (2008) Arsenic accumulation in rice (*Oryza sativa* L.): Human exposure through food chain. Ecotoxicol Environ Saf. doi: 10.1016/j.ecoenv.2007.01.005

Babu RC (2010) Breeding for drought resistance in rice: an integrated view from physiology to genomics. Electron J Plant Breed

Bajpai R, Upreti DK (2012) Accumulation and toxic effect of arsenic and other heavy metals in a contaminated area of West Bengal, India, in the lichen Pyxine cocoes (Sw.) Nyl. Ecotoxicol Environ Saf. doi: 10.1016/j.ecoenv.2012.06.001

Bak G, Lee E-J, Lee Y, et al (2013) Rapid Structural Changes and Acidification of Guard Cell Vacuoles during Stomatal Closure Require Phosphatidylinositol 3,5-Bisphosphate. Plant Cell 25:2202–2216

Bakhat HF, Zia Z, Fahad S, et al (2017) Arsenic uptake, accumulation and toxicity in rice plants: Possible remedies for its detoxification: A review. Environ Sci Pollut Res. doi: 10.1007/s11356-017-8462-2

Bandillo N, Raghavan C, Muyco PA, et al (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: Progress and potential for genetics research and breeding. Rice. doi: 10.1186/1939-8433-6-11

Banerjee M, Banerjee N, Bhattacharjee P, et al (2013) High arsenic in rice is associated with elevated genotoxic effects in humans. Sci Rep. doi: 10.1038/srep02195

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics. doi: 10.1093/bioinformatics/bth457

Bastías JM, Beldarrain T (2016) Arsenic translocation in rice cultivation and its implication for human health. Chil J Agric Res. doi: 10.4067/s0718-58392016000100016

Batista BL, Souza JMO, De Souza SS, Barbosa F (2011) Speciation of arsenic in rice and estimation of daily intake of different arsenic species by Brazilians through rice consumption. J Hazard Mater. doi: 10.1016/j.jhazmat.2011.04.087

Bhattacharya P, Samal AC, Majumdar J, Santra SC (2010) Accumulation of arsenic and its distribution in rice plant (*Oryza sativa* L.) in Gangetic West Bengal, India. Paddy Water Environ. doi: 10.1007/s10333-009-0180-z

BIG Data Center Members (2018) Database Resources of the BIG Data Center in 2018. Nucleic Acids Res. doi: 10.1093/nar/gkx897

Bouman B, Barker R, Humphreys E, et al (2013) Rice: Feeding the billions. In: Water for Food Water for Life: A Comprehensive Assessment of Water Management in Agriculture

Brady NC (2012) Soil Factors that Influence Rice Production. In: Proceedings of Symposium on Paddy Soils

Brammer H, Ravenscroft P (2009) Arsenic in groundwater: A threat to sustainable agriculture in South and South-east Asia. Environ. Int.

Brunetti P, Zanella L, De Paolis A, et al (2015) Cadmium-inducible expression of the ABC-type transporter AtABCC3 increases phytochelatin-mediated cadmium tolerance in Arabidopsis. J Exp Bot. doi: 10.1093/jxb/erv185

Caicedo AL (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. P Lo S Genet 3:1745–1756

Caldwell BK, Caldwell JC, Mitra SN, Smith W (2003) Searching for an optimum solution to the Bangladesh arsenic crisis. Soc Sci Med. doi: 10.1016/S0277-9536(02)00203-4

Carey A-M, Scheckel KG, Lombi E, et al (2010) Grain Unloading of Arsenic Species in Rice. Plant Physiol 152:309–319. doi: 10.1104/pp.109.146126

Chaou H, H.G. S, Glodek A, J. S (1998) Lucy-A Sequence Cleanup Program. In: Proceedings of the Tenth Annual Genome Sequencing and Annotation

Conference (GSACX), Miami, Florida

Clemens S (2006) Toxic metal accumulation, responses to exposure and mechanisms of tolerance in plants. Biochimie

Coleman JOD, Blake-Kalff MMA, Davies TGE (1997) Detoxification of xenobiotics by plants: Chemical modification and vacuolar compartmentation. Trends Plant Sci. 2:144–151

Cui Y, Zhang F, Xu J, et al (2015) Mapping quantitative trait loci in selected breeding populations: A segregation distortion approach. Heredity (Edinb). doi: 10.1038/hdy.2015.56

Dasgupta T, Hossain SA, Meharg AA, Price AH (2004) An arsenate tolerance gene on chromosome 6 of rice. New Phytol 163:45–49

De Abreu-Neto JB, Turchetto-Zolet AC, De Oliveira LFV, et al (2013) Heavy metal-associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. FEBS J 280:1604–1616

Dipankar C, Sengupta MK, Rahman MM, et al (2004) Groundwater arsenic contamination and its health effects in the Ganga-Meghna-Brahmaputra plain. J Environ Monit

Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. Genetics. doi: 10.1111/j.1369-7625.2010.00632.x

Duan G-L (2005) Characterization of Arsenate Reductase in the Extract of Roots and Fronds of Chinese Brake Fern, an Arsenic Hyperaccumulator. PLANT Physiol. doi: 10.1104/pp.104.057422

Duan G, Shao G, Tang Z, et al (2017) Genotypic and Environmental Variations in Grain Cadmium and Arsenic Concentrations Among a Panel of High Yielding Rice Cultivars. Rice. doi: 10.1186/s12284-017-0149-2

Eizenga G, McClung A, Mccouch S, et al (2012) Unraveling the rich phenotypic and genetic diversity in rice for varietal improvement. In: Proceedings of the 12th SABRAO Congress on Plant Breeding towards 2025

Ellis DR, Gumaelius L, Indriolo E, et al (2006) A Novel Arsenate Reductase from the Arsenic Hyperaccumulating Fern Pteris vittata. PLANT Physiol. doi: 10.1104/pp.106.084079

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379. doi: 10.1371/journal.pone.0019379

Environmental Protection Agency USEPA (1998) Method 7010: Graphite Furnace Atomic Absorption Spectrophotometry. US Environ Prot Agency. doi: 10.1104/pp.125.4.2154

Ewing B, Green P (1998) Base-Calling of Automated Sequencer Traces Using Phred.

II. Error Probabilities. Genome Res 8:186–194. doi: 10.1101/gr.8.3.186

FAO (2009) How to Feed the World in 2050. In: The High-Level Expert Forum on How to Feed the World in 2050

Feltus FA, Wan J, Schulze SR, et al (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14:1812–1819

Finatto T, de Oliveira AC, Chaparro C, et al (2015) Abiotic stress and genome dynamics: specific genes and transposable elements response to iron excess in rice. Rice. doi: 10.1186/s12284-015-0045-6

Frei M, Tetteh RN, Razafindrazaka AL, et al (2016) Responses of rice to chronic and acute iron toxicity: genotypic differences and biofortification aspects. Plant Soil 408:149–161. doi: 10.1007/s11104-016-2918-x

Fukao T, Bailey-Serres J (2008) Submergence tolerance conferred by Sub1A is mediated by SLR1 and SLRL1 restriction of gibberellin responses in rice. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.0807821105

Gao JP, Chao DY, Lin HX (2007) Understanding abiotic stress tolerance mechanisms: Recent studies on stress response in rice. J. Integr. Plant Biol.

Garai R, Chakraborty AK, Dey SB, Saha KC (1984) Chronic arsenic poisoning from tube-well water. J Indian Med Assoc

Ge S, Sang T, Lu B-R, Hong D-Y (2002) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci. doi: 10.1073/pnas.96.25.14400

Global Rice Science Partnership (2013) Rice almanac. International Rice Research Insttitute, Philippines

Gnanamanickam SS, Gnanamanickam SS (2009) Rice and Its Importance to Human Life. In: Biological Control of Rice Diseases

Gorny J, Billon G, Lesven L, et al (2015) Arsenic behavior in river sediments under redox gradient: A review. Sci. Total Environ.

Gregorio GB, Senadhira D, Mendoza RD, et al (2002) Progress in breeding for salinity tolerance and associated abiotic stresses in rice. F Crop Res. doi: 10.1016/S0378-4290(02)00031-X

Gregorio GB, Senadhira D, Mendoza RD (1997) Screening Rice for Salinity Tolerance. IRRI

Guan YS, Serraj R, Liu SH, et al (2010) Simultaneously improving yield under drought stress and non-stress conditions: A case study of rice (Oryza sativa L.). J Exp Bot 61:4145–4156

Guha Mazumder DN (2008) Chronic arsenic toxicity & human health. Indian J. Med.

Res.

Halim MA, Ghosh M, Nigar M, et al (2016) Screening of arsenic tolerance in rice at germination and early seedling stage as influenced by sodium arsenate. Jahangirnagar Univ J Biol Sci. doi: 10.3329/jujbs.v3i1.28273

Harper AL, Trick M, Higgins J, et al (2012) Associative transcriptomics of traits in the polyploid crop species Brassica napus. Nat Biotech 30:798–802

Hossain MF (2006) Arsenic contamination in Bangladesh - An overview. Agric. Ecosyst. Environ.

Hu Y (2014) Deep re-sequencing of a widely used maintainer line of hybrid rice for discovery of DNA polymorphisms and evaluation of genetic diversity. Mol Genet Genomics 289:303–315

Huang X, Kurata N, Wei X, et al (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature. doi: 10.1038/nature11532

Hughes MF, Beck BD, Chen Y, et al (2011) Arsenic exposure and toxicology: A historical perspective. Toxicol Sci. doi: 10.1093/toxsci/kfr184

Ishimaru Y, Masuda H, Suzuki M, et al (2007) Overexpression of the OsZIP4 zinc transporter confers disarrangement of zinc distribution in rice plants. J Exp Bot 58:2909–2915

Ishimaru Y, Suzuki M, Kobayashi T, et al (2005) OsZIP4, a novel zinc-regulated zinc transporter in rice. J Exp Bot 56:3207–3214

Islam MS, Thyssen GN, Jenkins JN, Fang DD (2015) Detection, Validation, and Application of Genotyping-by-Sequencing Based Single Nucleotide Polymorphisms in Upland Cotton. Plant Genome 8:. doi: 10.3835/plantgenome2014.07.0034

Islam S, Rahman MM, Islam MR, Naidu R (2016) Arsenic accumulation in rice: Consequences of rice genotypes and management practices to reduce human health risk. Environ. Int.

Islam S, Rahman MM, Rahman MA, Naidu R (2017) Inorganic arsenic in rice and rice-based diets: Health risk assessment. Food Control. doi: 10.1016/j.foodcont.2017.06.030

Jain M, Moharana KC, Shankar R, et al (2014) Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant Biotechnol J 12:253–264. doi: 10.1111/pbi.12133

Jeandroz S, Lamotte O (2017) Editorial: Plant Responses to Biotic and Abiotic Stresses: Lessons from Cell Signaling. Front Plant Sci. doi: 10.3389/fpls.2017.01772

Jernigan KL, Godoy J V., Huang M, et al (2018) Genetic Dissection of End-Use Quality

Traits in Adapted Soft White Winter Wheat. Front Plant Sci. doi: 10.3389/fpls.2018.00271

Jewel ZA, Ali J, Mahender A, et al (2019) Identification of Quantitative Trait Loci Associated with Nutrient Use Efficiency Traits, Using SNP Markers in an Early Backcross Population of Rice (*Oryza sativa* L.). Int J Mol Sci 20:. doi: 10.3390/ijms20040900

Jiang JQ, Ashekuzzaman SM, Jiang A, et al (2013) Arsenic contaminated groundwater and its treatment options in bangladesh. Int. J. Environ. Res. Public Health

Jung KH, Gho HJ, Giong HK, et al (2013) Genome-wide identification and analysis of Japonica and Indica cultivar-preferred transcripts in rice using 983 Affymetrix array data. Rice 6:1–13. doi: 10.1186/1939-8433-6-1

Khan MA, Islam MR, Panaullah GM, et al (2010) Accumulation of arsenic in soil and rice under wetland condition in Bangladesh. Plant Soil. doi: 10.1007/s11104-010-0340-3

Khush GS (2011) Origin, dispersal, cultivation and variation of rice. In: Oryza: From Molecule to Plant

Kinniburgh DG, Smedley PL (2001) Arsenic contamination of groundwater in Bangladesh. Br Geol Surv Tech Rep WC/00/19

Klein M, Burla B, Martinoia E (2006) The multidrug resistance-associated protein (MRP/ABCC) subfamily of ATP-binding cassette transporters in plants. FEBS Lett. 580:1112–1122

Krebs M, Beyhl D, Gorlich E, et al (2010) Arabidopsis V-ATPase activity at the tonoplast is required for efficient nutrient storage but not for sodium accumulation. Proc Natl Acad Sci. doi: 10.1073/pnas.0913035107

Kuramata M, Abe T, Kawasaki A, et al (2013) Genetic diversity of arsenic accumulation in rice and QTL analysis of methylated arsenic in rice grains. Rice 6:1–10. doi: 10.1186/1939-8433-6-1

Kuramata M, Abe T, Matsumoto S, Ishikawa S (2011) Arsenic accumulation and speciation in Japanese paddy rice cultivars. Soil Sci Plant Nutr. doi: 10.1080/00380768.2011.565479

Lafitte HR, Ismail A, Bennett J (2004) Abiotic stress tolerance in rice for Asia : progress and the future. *Proceedings 4th Int Crop Sci*

Lawler SP, Dritz DA (2005) Straw and winter flooding benefit mosquitoes and other insects in a rice agroecosystem. Ecol Appl. doi: 10.1890/03-5420

Lee JJ, Jang CS, Wang SW, et al (2008) Delineation of spatial redox zones using discriminant analysis and geochemical modelling in arsenic-affected alluvial aquifers. Hydrol Process. doi: 10.1002/hyp.6884

Lee S, Kim Y-Y, Lee Y, An G (2007) Rice P1B-Type Heavy-Metal ATPase, OsHMA9,

Is a Metal Efflux Protein. PLANT Physiol 145:831–842

Leiboff S, Li X, Hu H-C, et al (2015) Genetic control of morphometric diversity in the maize shoot apical meristem. Nat Commun 6:8974. doi: 10.1038/ncomms9974

Li F, Fan G, Wang K, et al (2014) Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet 46:567–572. doi: 10.1038/ng.2987

Li G, Sun GX, Williams PN, et al (2011) Inorganic arsenic in Chinese food and its cancer risk. Environ Int. doi: 10.1016/j.envint.2011.05.007

Li H, Vikram P, Singh RP, et al (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. BMC Genomics 16:216. doi: 10.1186/s12864-015-1424-5

Li S, Chou H-H (2004) LUCY2: an interactive DNA sequence quality trimming and vector removal tool. Bioinformatics 20:2865–2866. doi: 10.1093/bioinformatics/bth302

Li Z-K, Xu J-L (2007) Breeding for drought and salt tolerant rice (Oryza sativa L.): progress and perspectives. In: Advances in molecular breeding toward drought and salt tolerant crops. Springer, pp 531–564

Li Z, Ali J (2017) Breeding green super rice (GSR) varieties for sustainable rice cultivation. 1–22. doi: 10.19103/AS.2016.0003.05

Liao G, Wu Q, Feng R, et al (2016) Efficiency evaluation for remediating paddy soil contaminated with cadmium and arsenic using water management, variety screening and foliage dressing technologies. J Environ Manage. doi: 10.1016/j.jenvman.2016.01.008

Lin SS, Liao PB (1979) Evaluation of an extended aeration process for skokomish salmon processing wastewater treatment. Env Prot Technol Ser EPA. doi: 10.1104/pp.107.102236

Linares OF (2002) African rice (*Oryza glaberrima*): History and future potential. Proc Natl Acad Sci. doi: 10.1073/pnas.252604599

Liu T, Zeng J, Xia K, et al (2012) Evolutionary expansion and functional diversification of oligopeptide transporter gene family in rice. Rice 5:12

Liu X, Chen S, Chen M, et al (2019) Association Study Reveals Genetic Loci Responsible for Arsenic, Cadmium and Lead Accumulation in Rice Grain in Contaminated Farmlands. Front Plant Sci. doi: 10.3389/fpls.2019.00061

Lu Y, Dong F, Deacon C, et al (2010) Arsenic accumulation and phosphorus status in two rice (Oryza sativa L.) cultivars surveyed from fields in South China. Environ Pollut. doi: 10.1016/j.envpol.2009.12.022

Lubkowitz M (2011) The oligopeptide transporters: A small gene family with a diverse group of substrates and functions? Mol. Plant 4:407–415

Lucas SJ, Salantur A, Yazar S, Budak H (2017) High-throughput SNP genotyping of modern and wild emmer wheat for yield and root morphology using a combined association and linkage analysis. Funct Integr Genomics 17:667–685. doi: 10.1007/s10142-017-0563-y

Ma JF, Yamaji N, Mitani N, et al (2008) Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. Proc Natl Acad Sci 105:9931–9935

Maclean, J.L. et al. (2013) Rice almanac. 3rd edn, International Rice Research Institute, Philippines.

Mandal BK, Suzuki KT (2002) Arsenic round the world: a review.Talanta, 58 201-235

Marcaida M, Li T, Angeles O, et al (2014) Biomass accumulation and partitioning of newly developed Green Super Rice (GSR) cultivars under drought stress during the reproductive stage. F Crop Res 162:30–38. doi: 10.1016/j.fcr.2014.03.013

Mazumder DNG (2013) Human health effects of chronic arsenic toxicity. In: Arsenic: Sources, Environmental Impact, Toxicity and Human Health - A Medical Geology Perspective

McCough SR (1997) Report on QTL nomenclature. Rice Genet Newsl 14:11–13

McCough SR, Doerge RW (1995) QTL mapping in rice. Trends Genet. doi: 10.1016/S0168-9525(00)89157-X

Meharg AA, Hartley-Whitaker J (2002) Arsenic uptake and metabolism in arsenic resistant and nonresistant plant species. New Phytol.

Meharg AA, Zhao F-J (2012) Arsenic & Rice. Springer Netherlands, Dordrecht

Mehra P, Pandey BK, Giri J (2015) Genome-wide DNA polymorphisms in low Phosphate tolerant and sensitive rice genotypes. Sci Rep 5:13090. doi: 10.1038/srep13090

Meyer RS, Choi JY, Sanches M, et al (2016) Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat Genet. doi: 10.1038/ng.3633

Mitra A, Chatterjee S, Moogouei R, Gupta D (2017) Arsenic Accumulation in Rice and Probable Mitigation Approaches: A Review. Agronomy. doi: 10.3390/agronomy7040067

Mohan D, Pittman CU (2007) Arsenic removal from water/wastewater using adsorbents-A critical review. J. Hazard. Mater.

Molina J, Sikora M, Garud N, et al (2011) Molecular evidence for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci U S A. doi: 10.1073/pnas.1104686108

Muhammad S, Tahir Shah M, Khan S (2010) Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques

in Kohistan region, northern Pakistan. Food Chem Toxicol. doi: 10.1016/j.fct.2010.07.018

Muthayya S, Sugimoto JD, Montgomery S, Maberly GF (2014) An overview of global rice production, supply, trade, and consumption. Ann N Y Acad Sci. doi: 10.1111/nyas.12540

Nayar NM (2010) Origin of African rice from Asian rice. Second Africa Rice Congr. doi: http://dx.doi.org/10.1016/B978-0-12-417177-0.00005-X

Ng JC (2005) Environmental contamination of arsenic and its toxicological impact on humans. Environ. Chem.

Ni JJ, Wu P, Senadhira D, Huang N (1998) Mapping QTLs for phosphorus deficiency tolerance in rice (Oryza sativa L.). Theor Appl Genet. doi: 10.1007/s001220051030

Nickson R, McArthur J, Burgess W, et al (1998) Arsenic poisoning of Bangladesh groundwater [7]. Nature

Nickson RT, Mcarthur JM, Ravenscroft P, et al (2000) Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. Appl Geochemistry. doi: 10.1016/S0883-2927(99)00086-4

Njie M, Reed JD (1995) Potential of crop residues and agricultural by-products for feeding sheep in a Gambian village. Anim Feed Sci Technol. doi: 10.1016/0377-8401(94)00710-Q

Nordborg M, Hu TT, Ishino Y, et al (2005) The pattern of polymorphism in Arabidopsis thaliana. PLoS Biol 3:e196

Norton GJ, Deacon CM, Xiong L, et al (2010) Genetic mapping of the rice ionome in leaves and grain: Identification of QTLs for 17 elements including arsenic, cadmium, iron and selenium. Plant Soil 329:139–153

Norton GJ, Pinson SRM, Alexander J, et al (2012) Variation in grain arsenic assessed in a diverse panel of rice (Oryza sativa) grown in multiple sites. New Phytol. doi: 10.1111/j.1469-8137.2011.03983.x

Nriagu JO, Bhattacharya P, Mukherjee AB, et al (2007) Arsenic in soil and groundwater: an overview. Trace Met. other Contam. Environ.

O. Akinbil C, Haque AMM (2012) Arsenic Contamination in Irrigation Water for Rice Production in Bangladesh: A Review. Trends Appl Sci Res. doi: 10.3923/tasr.2012.331.349

Ott A, Liu S, Schnable JC, et al (2017) tGBS®genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Res. doi: 10.1093/nar/gkx853

Ouyang S, Zhu W, Hamilton J, et al (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. Nucleic Acids Res 35:8–11

Pan X, Zhang Q, Yan W, et al (2012) Development of Genetic Markers Linked to Straighthead Resistance through Fine Mapping in Rice (Oryza sativa L.). PLoS One. doi: 10.1371/journal.pone.0052540

Pandey S, Rai R, Rai LC (2015) Biochemical and Molecular Basis of Arsenic Toxicity and Tolerance in Microbes and Plants. Elsevier Inc.

Pandey V, Shukla A (2015) Acclimation and Tolerance Strategies of Rice under Drought Stress. Rice Sci.

Pang Y, Chen K, Wang X, et al (2017) Simultaneous Improvement and Genetic Dissection of Salt Tolerance of Rice (Oryza sativa L.) by Designed QTL Pyramiding. Front Plant Sci 8:1–11. doi: 10.3389/fpls.2017.01275

Paradis E, Gosselin T, Goudet J, et al (2017) Linking genomics and population genetics with R. In: Molecular Ecology Resources

Park J, Song WY, Ko D, et al (2012) The phytochelatin transporters AtABCC1 and AtABCC2 mediate tolerance to cadmium and mercury. Plant J 69:278–288

Pascual L, Albert E, Sauvage C, et al (2016) Dissecting quantitative trait variation in the resequencing era: Complementarity of bi-parental, multi-parental and association panels. Plant Sci 242:120–130

Patrick S. Schnable tGBS: A Next Generation Genotyping-by-Sequencing Technology. In: Plant and Animal Genome XXII Conference

Paumi CM, Chuk M, Snider J, et al (2009) ABC Transporters in Saccharomyces cerevisiae and Their Interactors: New Technology Advances the Biology of the ABCC (MRP) Subfamily. Microbiol Mol Biol Rev 73:577–593

Perata P, Voesenek LACJ (2007) Submergence tolerance in rice requires Sub1A, an ethylene-response-factor-like gene. Trends Plant Sci.

Pezeshki SR, DeLaune RD (2012) Soil Oxidation-Reduction in Wetlands and Its Impact on Plant Functioning. Biology (Basel). doi: 10.3390/biology1020196

Pillai TR, Yan W, Agrama HA, et al (2010) Total grain-arsenic and arsenic-species concentrations in diverse rice cultivars under flooded conditions. Crop Sci. doi: 10.2135/cropsci2009.10.0568

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. PLoS One 7:e32253. doi: 10.1371/journal.pone.0032253

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J 5:92. doi: 10.3835/plantgenome2012.05.0005

Puri A, Kumar M (2012) A review of permissible limits of drinking water. Indian J Occup Environ Med. doi: 10.4103/0019-5278.99696

Purugganan MD (2014) An evolutionary genomic tale of two rice species. Nat Genet.

doi: 10.1038/ng.3071

R Core Team (2015) R: A Language and Environment for Statistical Computing

Raab A, Williams PN, Meharg A, Feldmann J (2007) Uptake and translocation of inorganic and methylated arsenic species by plants. Environ Chem. doi: 10.1071/EN06079

Raghavan C, Mauleon R, Lacorte V, et al (2019) Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population. G3: Genes|Genomes|Genetics. doi: 10.1534/g3.117.042101

Rahman MA, Hasegawa H, Rahman MM, et al (2008) Straighthead disease of rice (Oryza sativa L.) induced by arsenic toxicity. Environ Exp Bot. doi: 10.1016/j.envexpbot.2007.07.016

Rahman MA, Hasegawa H, Rahman MM, et al (2007) Accumulation of arsenic in tissues of rice plant (Oryza sativa L.) and its distribution in fractions of rice grain. Chemosphere. doi: 10.1016/j.chemosphere.2007.05.044

Rahman MA, Rahman MM, Hasegawa H (2012) Arsenic-induced straighthead: An impending threat to sustainable rice production in South and South-East Asia! Bull Environ Contam Toxicol. doi: 10.1007/s00128-011-0490-x

Rahman MM, Naidu R, Bhattacharya P (2009) Arsenic contamination in groundwater in the Southeast Asia region. Environ. Geochem. Health

Rao CS, Lal R, Prasad JVNS, et al (2015) Potential and challenges of rainfed farming in India. Adv Agron. doi: 10.1016/bs.agron.2015.05.004

Reimann C, Garrett RG (2005) Geochemical background - Concept and reality. Sci Total Environ. doi: 10.1016/j.scitotenv.2005.01.047

Richards M, Sander BO (2014) Alternate wetting and drying in irrigated rice. J AHIMA. doi: 10.1016/j.techfore.2006.05.021

Roy P, Saha A (2002) Metabolism and toxicity of arsenic: A human carcinogen. Curr. Sci.

Roychowdhury T (2008) Impact of sedimentary arsenic through irrigated groundwater on soil, plant, crops and human continuum from Bengal delta: Special reference to raw and cooked rice. Food Chem Toxicol. doi: 10.1016/j.fct.2008.05.019

Sánchez-Bermejo E, Castrillo G, Del Llano B, et al (2014) Natural variation in arsenate tolerance identifies an arsenate reductase in Arabidopsis thaliana. Nat Commun. doi: 10.1038/ncomms5617

Schnable PS, Liu. S, Wu W (2013) GENOTYPING BY NEXT-GENERATION SEQUENCING. United States Pat. Appl. Publ.

Shamsi IH, Jilani G, Zhang GP, Kang W (2008) Cadmium stress tolerance through potassium nutrition in soybean. Asian J Chem 20:1099–1108

Shrestha A, Dziwornu AK, Ueda Y, et al (2018) Genome-wide association study to identify candidate loci and genes for Mn toxicity tolerance in rice. PLoS One 13:1–15

Shri M, Kumar S, Chakrabarty D, et al (2009) Effect of arsenic on growth, oxidative stress, and antioxidant system in rice seedlings. Ecotoxicol Environ Saf 72:1102–1110

Singh Chauhan B, Opeña J, Ali J (2015) Response of 10 Elite "Green Super Rice" Genotypes to Weed Infestation in Aerobic Rice Systems. Plant Prod Sci 18:228–233. doi: 10.1626/pps.18.228

Smith AH, Lingas EO, Rahman M (2000) Contamination of drinking-water by arsenic in Bangladesh: A public health emergency. Bull World Health Organ

Sohn E (2014) The toxic side of rice. Nature Outlook:5–6. doi: 10.1038/514S62a

Song W-Y, Park J, Mendoza-Cozatl DG, et al (2010) Arsenic tolerance in Arabidopsis is mediated by two ABCC-type phytochelatin transporters. Proc Natl Acad Sci. doi: 10.1073/pnas.1013964107

Stankovic I (2015) Codex Alimentarius. In: Encyclopedia of Food and Health

Sturchio E, Zanellato M, Minoia C, Bemporad E (2013) Arsenic: Environmental contamination and exposure. In: Arsenic: Sources, Environmental Impact, Toxicity and Human Health - A Medical Geology Perspective

Su C, Jiang L, Zhang W (2014) A review on heavy metal contamination in the soil worldwide: Situation, impact and remediation techniques. Environ Skept Critics. doi: 10.1037/a0036071

Subbaiyan GK (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. Plant Biotechnol J 10:623–634

Sultana F (2013) Water, technology, and development: Transformations of development technonatures in changing waterscapes. Environ Plan D Soc Sp. doi: 10.1068/d20010

Suriyagoda LDB, Dittert K, Lambers H (2018a) Mechanism of arsenic uptake, translocation and plant resistance to accumulate arsenic in rice grains. Agric. Ecosyst. Environ.

Suriyagoda LDB, Dittert K, Lambers H (2018b) Arsenic in Rice Soils and Potential Agronomic Mitigation Strategies to Reduce Arsenic Bioavailability: A Review. Pedosphere 28:363–382

Svedberg EB, Mallett JJ, Bendersky LA, et al (2006) A Structural Study of Electrodeposited Fe on n-GaAs(001). J Electrochem Soc. doi: 10.1149/1.2353782

Talukder ASMHM, Meisner CA, Sarkar MAR, Islam MS (2011) Effect of water management, tillage options and phosphorus status on arsenic uptake in rice.

Ecotoxicol Environ Saf. doi: 10.1016/j.ecoenv.2010.11.004

Tang RJ, Liu H, Yang Y, et al (2012) Tonoplast calcium sensors CBL2 and CBL3 control plant growth and ion homeostasis through regulating V-ATPase activity in Arabidopsis. Cell Res 22:1650–1665

Tareq SM, Islam SMN, Rahmam MM, Chowdhury DA (2010) Arsenic Pollution in Groundwater of Southeast Asia : an Overview on Mobilization Process and Health Effects. Bangladesh J Environ Res

Thomson MJ (2014) High-Throughput SNP Genotyping to Accelerate Crop Improvement. Plant Breed Biotechnol 2:195–212

The 3,000 rice genomes project (2014). Gigascience 3:7. doi:10.1186/2047-217X-3-7

Thomson MJ, Singh N, Dwiyanti MS, et al (2017) Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. Rice 10:. doi: 10.1186/s12284-017-0181-2

Tripathi P, Mishra A, Dwivedi S, et al (2012) Differential response of oxidative stress and thiol metabolism in contrasting rice genotypes for arsenic tolerance. Ecotoxicol Environ Saf. doi: 10.1016/j.ecoenv.2011.12.019

Tripathi P, Tripathi RD, Singh RP, et al (2013) Arsenite tolerance in rice (Oryza sativa L.) involves coordinated role of metabolic pathways of thiols and amino acids. Environ Sci Pollut Res 20:884–896. doi: 10.1007/s11356-012-1205-5

Tuli R, Chakrabarty D, Trivedi PK, Tripathi RD (2010) Recent advances in arsenic accumulation and metabolism in rice. Mol Breed. doi: 10.1007/s11032-010-9412-6

Tyczewska A, Woźniak E, Gracz J, et al (2018) Towards Food Security: Current State and Future Prospects of Agrobiotechnology. Trends Biotechnol.

Vahter M, Concha G (2010) Role of Metabolism in Arsenic Toxicity. Pharmacol Toxicol. doi: 10.1111/j.1600-0773.2001.890101.x

Wang M, Yu Y, Haberer G, et al (2014) The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nat Genet. doi: 10.1038/ng.3044

Wang S, Meyer E, McKay JK, Matz M V (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Meth 9:808–810

Wang W, Fu B, Ali J, et al (2015) Genome-Wide Responses to Selection and Genetic Networks Underlying Submergence Tolerance in Rice. Plant Genome 8:0. doi: 10.3835/plantgenome2014.10.0066

Wang WS, Pan YJ, Zhao XQ, et al (2011) Drought-induced site-specific DNA methylation and its association with drought tolerance in rice (Oryza sativa L.). J Exp Bot 62:1951–1960. doi: 10.1093/jxb/erq391

Wenzel WW, Alloway BJ (2013) Chapter 9 Arsenic. In: Heavy metals in Soils

WHO (2010) Exposure to Arsenic: A Major Public Health Concern. Agriculture 5. doi: 10.1016/j.ecoenv.2011.12.007

Wissuwa M, Kondo K, Fukuda T, et al (2015) Unmasking novel loci for internal phosphorus utilization efficiency in rice germplasm through genome-wide association analysis. PLoS One. doi: 10.1371/journal.pone.0124215

Wissuwa M, Wegner J, Ae N, Yano M (2002) Substitution mapping of Pup1: A major QTL increasing phosphorus uptake of rice from a phosphorus-deficient soil. Theor Appl Genet. doi: 10.1007/s00122-002-1051-9

Wong M, Dick R, Wu Q, Basta N (2012) Biomethylation of Arsenic in Contaminated Soils. In: Environmental Contamination

Wu C, Ye Z, Shu W, et al (2011) Arsenic accumulation and speciation in rice are affected by root aeration and variation of genotypes. J Exp Bot. doi: 10.1093/jxb/erq462

Wu L-B, Shhadi MY, Gregorio G, et al (2014) Genetic and physiological analysis of tolerance to acute iron toxicity in rice. Rice 7:8

Wu LB, Ueda Y, Lai SK, Frei M (2017) Shoot tolerance mechanisms to iron toxicity in rice (Oryza sativa L.). Plant Cell Environ 40:570–584

Wu LB, Ueda Y, Lai SK, Frei M (2016) Shoot tolerance mechanisms to iron toxicity in rice (Oryza sativa L.). Plant, Cell Environ. doi: 10.1111/pce.12733

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. doi: 10.1093/bioinformatics/btq057

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet. 125:201–210

Xia L, Zou D, Sang J, et al (2017) Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. J Genet Genomics 44:235–241

Xu K, Xu X, Fukao T, et al (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature. doi: 10.1038/nature04920

Xu X, Pan S, Cheng S, et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. doi: 10.1038/nature10158

Xue Y, Wan JM, Jiang L, et al (2006) QTL analysis of aluminum resistance in rice (Oryza sativa L.). Plant Soil. doi: 10.1007/s11104-006-9086-3

Yan W, Dilday RH, Tai TH, et al (2005) Differential response of rice germplasm to straighthead induced by arsenic. Crop Sci. doi: 10.2135/cropsci2004.0348

Yang Y, Zhang A, Chen Y, et al (2018) Impacts of silicon addition on arsenic

fractionation in soils and arsenic speciation in Panax notoginseng planted in soils contaminated with high levels of arsenic. Ecotoxicol Environ Saf. doi: 10.1016/j.ecoenv.2018.07.015

Yorobe JM, Ali J, Pede VO, et al (2016) Yield and income effects of rice varieties with tolerance of multiple abiotic stresses: The case of green super rice (GSR) and flooding in the Philippines. Agric Econ (United Kingdom) 47:261–271

Yu LJ, Luo YF, Liao B, et al (2012) Comparative transcriptome analysis of transporters, phytohormone and lipid metabolism pathways in response to arsenic stress in rice (Oryza sativa). New Phytol 195:97–112. doi: 10.1111/j.1469-8137.2012.04154.x

Zhang J, Zhu Y-G, Zeng D-L, et al (2007) Mapping quantitative trait loci associated with arsenic accumulation in rice (Oryza sativa). New Phytol 0:071107070910002-??? doi: 10.1111/j.1469-8137.2007.02267.x

Zhang J, Zhu YG, Zeng DL, et al (2008) Mapping quantitative trait loci associated with arsenic accumulation in rice (Oryza sativa). New Phytol 177:350–355. doi: 10.1111/j.1469-8137.2007.02267.x

Zhang M, Pinson SRM, Tarpley L, et al (2014) Mapping and validation of quantitative trait loci associated with concentrations of 16 elements in unmilled rice grain. Theor Appl Genet 127:137–165

Zhang YM, Gai J (2009) Methodologies for segregation analysis and QTL mapping in plants. Genetica. doi: 10.1007/s10709-008-9313-3

Zhao F-J, McGrath SP, Meharg AA (2010a) Arsenic as a Food Chain Contaminant: Mechanisms of Plant Uptake and Metabolism and Mitigation Strategies. Annu Rev Plant Biol 61:535–559

Zhao FJ, Ago Y, Mitani N, et al (2010b) The role of the rice aquaporin Lsi1 in arsenite efflux from roots. New Phytol 186:392–399

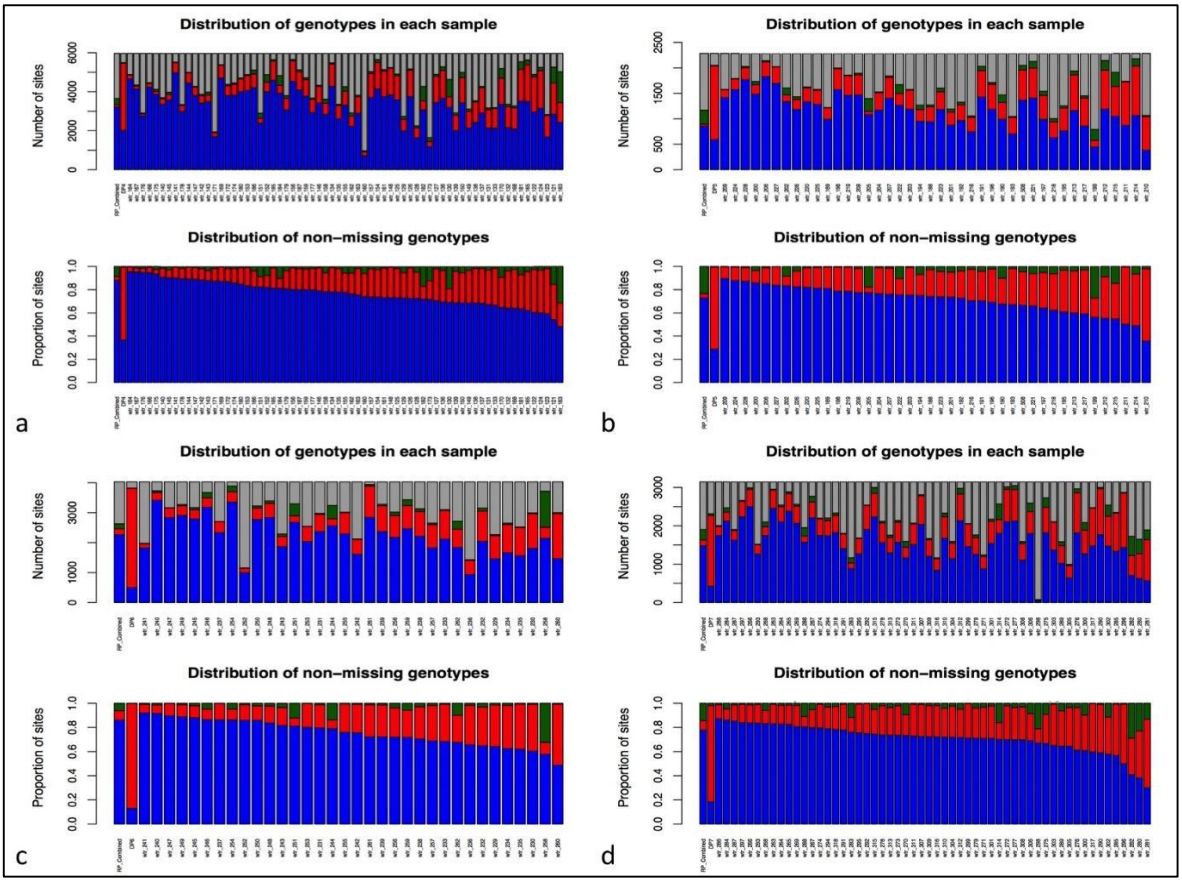
Zhao FJ, Ma JF, Meharg AA, McGrath SP (2009) Arsenic uptake and metabolism in plants. New Phytol. doi: 10.1111/j.1469-8137.2008.02716.x

Zhou Q, Teng Y, Liu Y (2017) A study on soil-environmental quality criteria and standards of arsenic. Appl Geochemistry 77:158–166

Zhu Y, Chen K, Mi X, et al (2015) Identification and Fine Mapping of a Stably Expressed QTL for Cold Tolerance at the Booting Stage Using an Interconnected Breeding Population in Rice. PLoS One 10:e0145704. doi: 10.1371/journal.pone.0145704

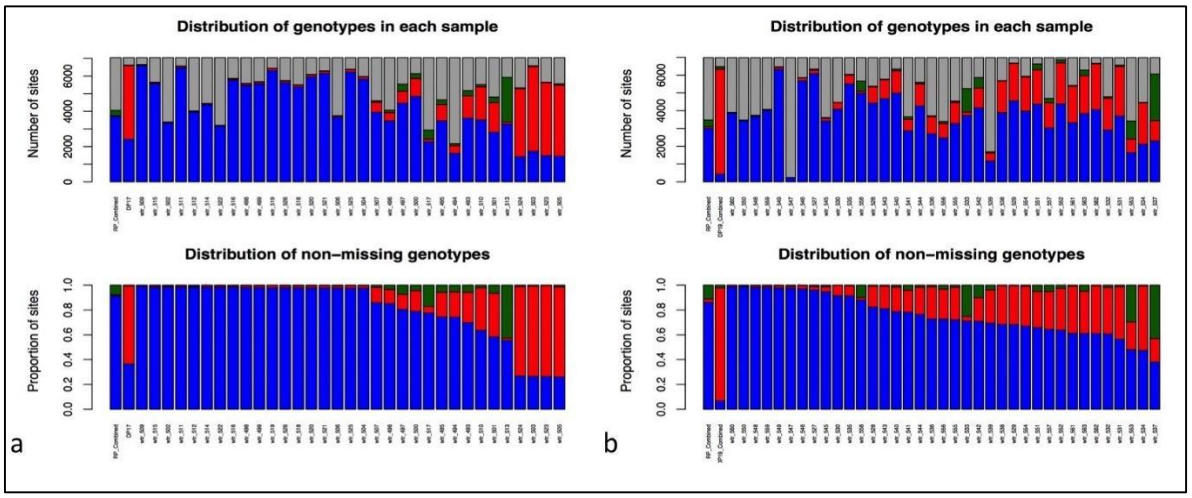
Zhu YG, Williams PN, Meharg AA (2008) Exposure to inorganic arsenic from rice: A global health issue? Environ Pollut. doi: 10.1016/j.envpol.2008.03.015

SUPPLEMENTARY DATA



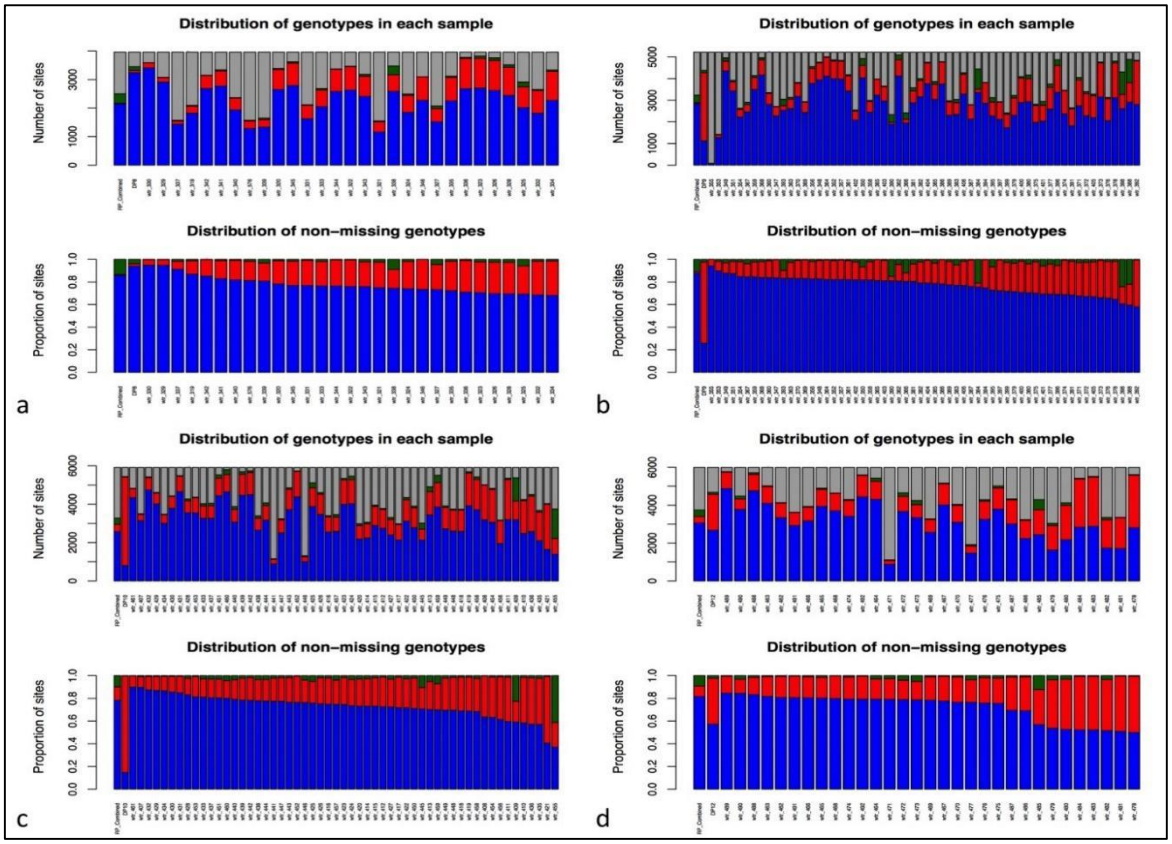
Supplementary Figure 2. 1. LMD50 SNPs genotype summary by sample

a) sub-population 2, b) sub-population 3, c) sub-population 4, d) sub-population 5



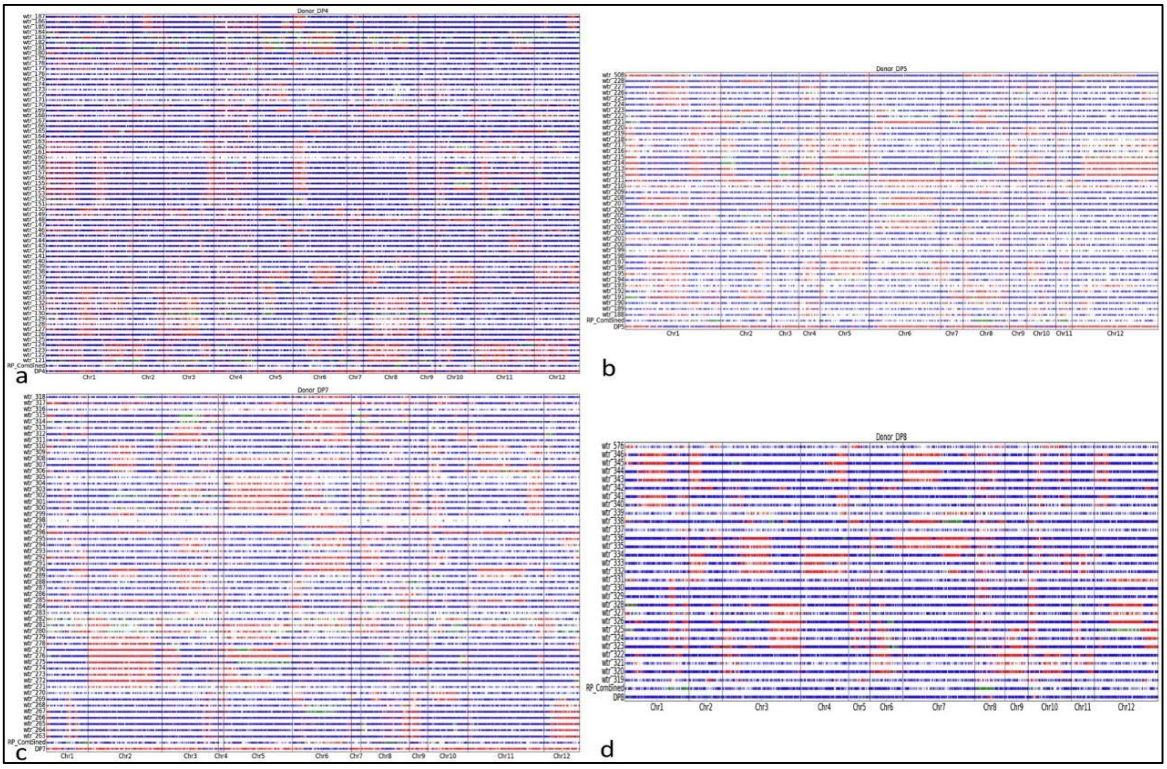
Supplementary Figure 2. 2. LMD50 SNPs genotype summary by sample

a) sub-population 10 and b) sub-population 11



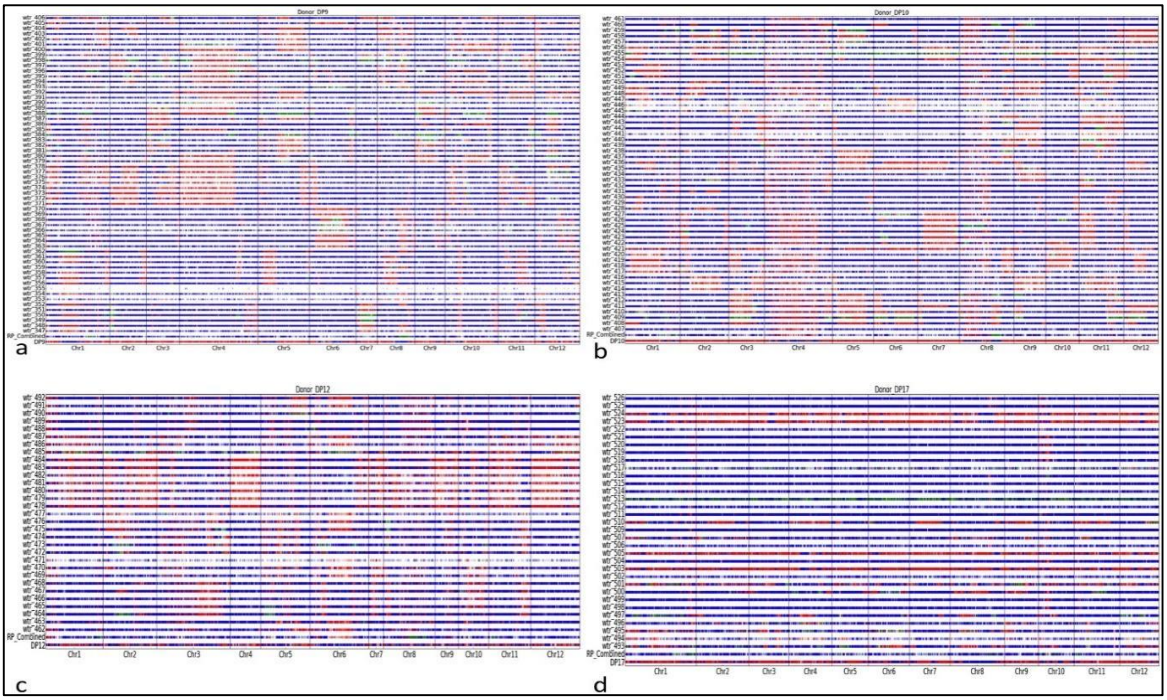
Supplementary Figure 2. 3. LMD50 SNPs genotype summary by sample

a) sub-population 6, b) sub-population 7, c) sub-population 8, d) sub-population 9



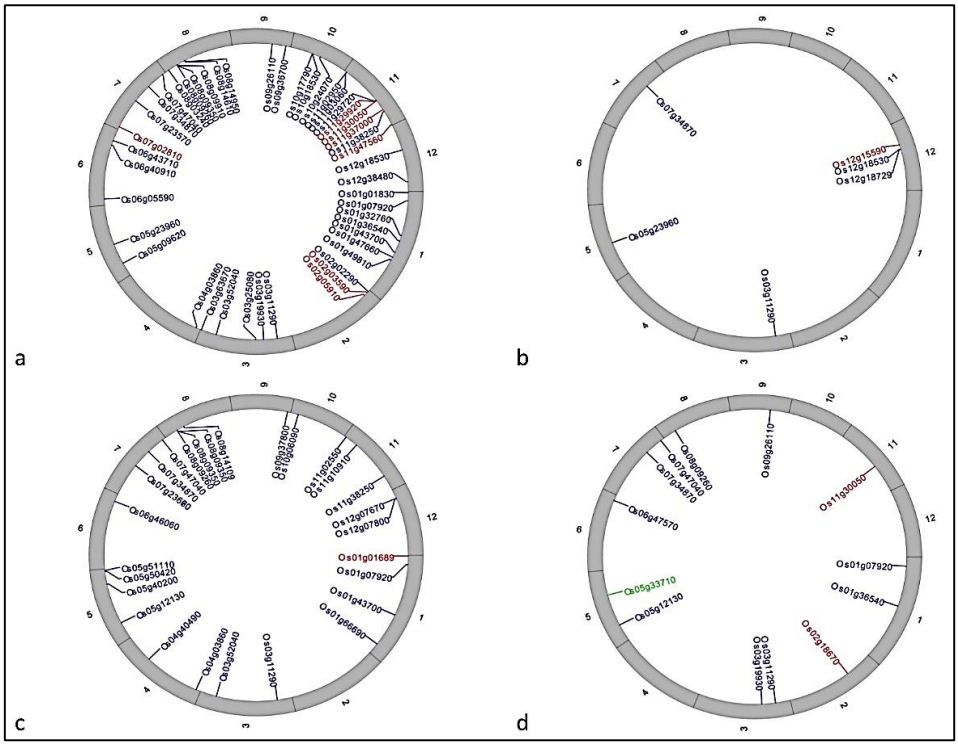
Supplementary Figure 2. 4. LMD50 SNPs representation on chromosomal basis.

a) sub-population 2, b) sub-population 3, c) sub-population 4, d) sub-population 5

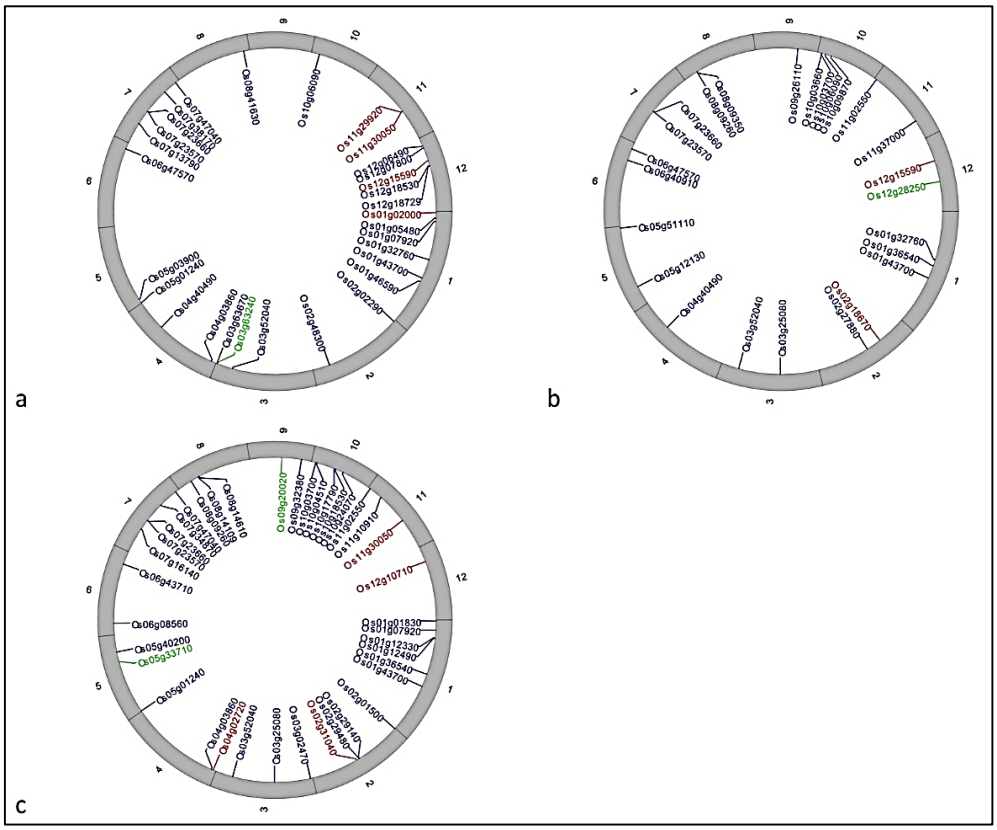


Supplementary Figure 2. 5. LMD50 SNPs representation on chromosomal basis a) sub-population 6, b) sub-population 7, c) sub-population 8, d) sub-population 9

Supplementary Figure 2. 6. LMD50 SNPs representation on chromosomal basis a) sub-population 10 and b) sub-population 11



Supplementary Figure 3. 1. Distribution of non-synonymous deleterious SNPs a) Sub-population 2, b) sub-population 3, c) sub-population 4, d) sub-population 5



Supplementary Figure 3. 2. Distribution of non-synonymous deleterious SNPs a) Sub-population 6, b) sub-population 7 and c) sub-population 8.

ACKNOWLEDGEMENTS

I would like to extend thanks to the many people, in many countries, who so generously to the work presented in this thesis.

I like to express my sincere thanks to Dr. Jauhar Ali, my research advisor and senior plant breeder at IRRI for providing me with the opportunity to work under him in the prestigious institute IRRI. I greatly benefited from his scientific discussion, his enthusiasm to tackle scientific questions and his ability to put sophisticated ideas into simple terms. I particularly appreciate his supportive nature towards his students.

Foremost, I would like to express my sincere gratitude and immense appreciation to Prof. Dr. Michael Frei, my doctoral supervisor and Professor at Rheinischen Friedrich-Wilhelms-Universität Bonn for providing me with the opportunity of a lifetime to work under him. I appreciate his supportive and friendly nature towards everyone. I admire his motivational attitude and his ability to finish work in a matter of seconds. I feel very privileged to have had the great chance to be supervised by him. I appreciate his support not only to my research but also to my entire stay in Germany.

Wish to express my heartfelt thanks to Dr. Zhikang Li, Director of Green Super Rice project for his generously helping in all possible ways to complete my research successfully and for his valuable guidance during my research.

My most profound appreciation and sincere gratefulness to Prof. Dr. Mathias Becker, for providing me with the opportunity to come to Germany for writing my thesis. I always had a great time talking with him regarding Arsenic contamination in rice. His vast network and scientific knowledge inspired my work and me.

I would also like to thank Prof. Dr. Jens Leon for being my second supervisor. I appreciate his support and helping nature.

My sincere thanks also go to Prof. Dr. Gabriel Schaaf for his helping tendency and constant encouragement throughout my stay in Germany.

I would like to register my sincere thanks to Dr. Lin-Bo Wu for his continual support. I admire his helping tendency, and very grateful for his help during my entire stay in Germany.

Special thanks to Mr.Bernhard Liese for his helping nature and everlasting support throughout my stay in the Philippines and Germany. He made my stay in Germany very pleasant and memorable.

I would like to express my sincere thanks and gratitude for the financial support from the Bill and Melinda Gates Foundation (BMGF), Department of Agriculture (DA) Philippines through the Green Super Rice project at the International Rice Research Institute (IRRI), Philippines and German Academic Exchange Service (DAAD) for providing scholarship to write my thesis in Germany.

Also, I express my sincere thanks to all the colleagues in the Green Super Rice Program, who directly or indirectly helped me in the completion of my PhD work.

I record my special thanks to Prof. Dr. A. S. Rao (retired), Prof. Dr. Chellam Balasundaram (retired), Dr. Michael Dingkuhn, Dr. Nickolai Alexandrov, Dr. Kenneth L. McNally, Dr. Reiner Wassman, Dr. Bjoern Ole Sander, Dr. James Quilty, Dr. Shalabh Dixit and Dr. Anshuman Singh for their valuable guidance and support throughout my entire carrier.

I like to express my sincere thanks to my office colleague Frau. Deborah Rupprecht for constant support and motivation in the office.

I also like to express my sincere thanks to Frau. Magdalena Gelep for making my stay pleasant in Germany.

It would be unfaithful if I fail to express my sincere gratitude to my friends who have been of immense help and moral support to me throughout my life. I express my thanks to Mr. Prakash, Mr. Arulvignesh, Dr. Dinesh, Mr. Venkadesh, Mr. Godwin, Dr. Sri Krishnan, Dr. Murali Kannan, Mr. Gowtham, Mr. Prabakaran, Dr. Janani, Ms. Roobini, Ms. Priyam Vadha, Ms. Monisha, Ms. Haley Sater, Ms. Liana Acevedo, Ms. Lisa Straussberger, Dr. Ruben Verbeek, Mr. Baldur Janz, Ms. Pia Schneider, Dr. Sarah Maarastawi, Dr. Irabella Fuhr, Mr. Justin Makiling, Ms. Marie Habito, Ms. Cecilia Fischer, Ms. Ilaria Doimo, Ms. Rachana Srinivasa, Ms. Rowely Parico, Dr. Thanai Joshi, Ms. Smitha Joshi, Ms. Amani Mahindawansha, Dr. Molly Shuman Goodier, Ms. Inga Gingembre, Dr. Kath Whitehouse, Ms. Noriko Edahiro, Ms. Feona Elize Gumabao, Dr. Ritu Bansal, Dr. Vishnu Vardhni, Dr. Umar Aslam, Ms. Muniba Fauze, Dr. Zilhas Ahmed Jewel, Ms. Andriele Warich, Ms. Louisa Sophie Wember, Ms. Isabel Cadavid, Mr. Dereje Demie, Mr. Kai Leers, Ms. Ambika Pandey, Mr.Christian Heck, Ms. Zhaleh Haghighat, Mr. Felix Holtkamp, Dr. MD Ashrafuzzaman and my entire friends.

It is beyond words if, I try to express my special thanks to my brother Prof. Dr. Jayaseelan Murugaiyan. Has been my role model right from my childhood. He was the one who laid the path for my career as a researcher and extended his full support even during his toughest times of life. His constant support and motivation helped me to overcome difficult situations. I feel very thankful for his everlasting support.

At last but definitely not the least, words become small to express my deep sense of indebtedness and affection to my father & mother for their infinite love, blessings and support for this achievement.