SIMPLE SEQUENCE REPEAT ANALYSIS OF 2015 WHEAT STEM RUST PATHOGEN (*Puccinia graminis* f. sp. *Tritici*); (*Pgt*) POPULATION IN KENYA.

EDGAR OTIENO OKELLO

A THESIS SUBMITTED TO GRADUATE SCHOOL IN PARTIAL FULFILMENT FOR THE REQUIREMENTS OF THE DEGREE OF MASTER OF SCIENCE IN PLANT PROTECTION OF UNIVERSITY OF ELDORET, KENYA

OCTOBER, 2019

DECLARATION

DECLARATION BY THE STUDENT

I hereby declare that this is my original work and has not been presented in this or any other university for the award of a degree.

Signature……Date…….

EDGAR OTIENO OKELLO

AGR/PGC/005/13

DECLARATION BY THE SUPERVISORS

This thesis has been submitted for examination with our approval as the candidate's supervisors.

PROF. JULIUS ONYANGO OCHUODHO

School of Agriculture and Biotechnology

University of Eldoret

Signature……Date……

PROF. LES J. SZABO

USDA Cereal Disease Laboratory (CDL)

University of Minnesota, USA

Signature……Date……

DEDICATION

This thesis is dedicated to my family for their patience and moral support.

"I have fought a good fight, I have finished [my] course, I have kept the faith": 2 Timothy 4:7

ABSTARCT

Wheat (*Triticum aestivium*) production in Kenya is severely hampered by emerging and re-emerging pathogens, stem rust Ug99 and related race group being one of the major constraints. The challenge has been largely attributed due to rapid evolution of races within the Ug99 lineage defeating deployed resistance genes. The objectives of this study were to determine the level of genetic variation of *Pgt* population in Kenya (2015) whether it is single or multiple and the driving force of the level of genetic variation whether due to asexual mutations, sexual recombination or migration forces. A total of 536 single uredinial-pustules were collected from wheat fields in the four-main wheat growing regions: Mount Kenya (101 samples), Central Rift (171 samples), North Rift (108 samples) and South Rift (156 samples). Out of the total samples collected, 104 single uredinial-pustule collections were successfully genotyped using 10 *Pgt* Simple Sequence Repeats (SSR) markers. Allele frequency distribution ranged from 2 to 4 per locus with an average of 3.10 per locus. Observed heterozygosity (H_o) for each of the ten SSR loci were significantly different ($P < 0.001$) than the expected heterozygosity (H_e) with Fixation Index (F) ranging from (-0.083 to -0.765). Five SSR Multi-locus Genotypes (SSR-MLGs) were identified. Neighbour joining (NJ) tree analysis clustered the five SSR-MLGs based on samples chosen from wheat growing fields with clade I (Ug99 race group) and clade IV-B (race TKTTF/TTTTF) indicating 2015 *Pgt* population being single. Analysis of molecular variance (AMOVA) showed that the majority of the variation occurred within the samples (91%) rather than between regions (9%). *Psex* values from each SSR-MLGs were significantly different ($P <0.001$) than the simulated *psex* (1.60E-08) meaning they likely belong to the same genet and clonal in nature. This would confirm that the SSR-MLGs generated are as a result of asexual mutation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/ACRONYMS

CTAB	Cetyl trimethylammonium bromide
DNA	Deoxyribonucleic acid
Genet	A dependent member of a clone originating from asexual reproduction of a single ancestor.
ITS	Internal transcribed spacer region of the nuclear ribosomal
KALRO	Kenya Agricultural and Livestock Research Organization
MLGs	Multi-locus genotypes
PCR	Polymerase chain reaction
Pathotype	Group of organism (of the same species) that have the same virulence
Pgt	*Puccinia graminis* Pers,: Per. f. sp. *tritici* Eriks
Ramet	An independent member of a clone originating vegetatively not sexually from a single ancestor.
Race phenotyping/typing	A procedure for virulence phenotyping of isolates of stem rust using spray inoculation on selected differential lines showing race-specificity
Sr	Stem rust resistance gene
SSR	Simple Sequence Repeats markers
SAMs	Spatial Analysis Methods
STMs	Sequence-Tagged Microsatellite sites
TTKSK	Acronym for Ug99 according to the North American race nomenclature
Ug99	Stem rust race discovered in Uganda on wheat in 1998

ACKNOWLEDGMENT

I would like to acknowledge Prof. Les J. Szabo and Prof Julius Ochuodho, my advisors, for their support, assistance and advice throughout this process. I am immensely grateful for the assistance and insightful discussions with members of USDA Cereal Disease Lab (St Paul Minnesota USA) most notably Dr. Jerry Johnson, Kim-Phuong Nguyen and Nathan Buermann. I would also like to appreciate the KALRO pathology team of R. Wanyera, H. Wanga, P. Kinyanjui and E. Githinji for their technical contributions. In addition to this I appreciate the Norman E. Borlaug Leadership Enhancement in Agriculture Program who funded my laboratory experiment through USAID. Lastly but not least I would like to thank the faculty, staff and fellow graduate students in the Department of Seed, Crop and Horticulture.

CHAPTER ONE

INTRODUCTION

1.0 Background Information

Stem rust caused by *Puccinia graminis* Pers,: Per. f. sp. *tritici* Eriks. and E. Henn (*Pgt*) causes significant losses to wheat (*Triticum aestivum*) production worldwide. It has been highlighted as a re-emerging infectious disease due to its increased incidence, geographic and host range, and finally its change in pathogenesis due to its constant evolution. These coupled characteristics make this pathogen become a threat to global food security and social stability (Vurro *et al*,. 2010). Majority of the affected countries depend on wheat as their staple crop and major source of calories accounting for up to 40% (Vurro *et al*,. 2010). If the challenge is not addressed at an early stage, the dependent countries will import more food due to low production of the same crop. In addition to this, the country's economy become imbalanced as the crop prices increases due to high demand and less supply.

The pathogen belongs to rust fungi (Phylum *Basidiomycota*, Order *Pucciniales*) one of the fungal plant pathogens that has elicited interested in the recent times due to its diversity. *Pgt* is one of the most destructive diseases of wheat and barley since it infects most parts of spikes, leaf blades, leaf sheath, glumes and stem (Roelfs *et al*,. 1992). The consequence of stem rust infection is the reduction of the nutrients and photosynthates to the developing kernels, thus resulting in shrivelled grain (Park, 2007). Deployment of cultivars carrying stem rust resistance (*Sr*) genes has effectively controlled stem rust over the past 50 years (Singh *et al*., 2011a).

In 1998, stem rust was observed on a breeding line containing the stem rust resistance gene *Sr*31 in Uganda. The following year, this isolate was characterized (*Pgt* race TTKSK) and

later designated as Ug99 (Pretorius *et al*., 2000). With the emergence of this new *Pgt* race most of the wheat germplasm has become susceptible (Singh *et al*,. 2011a). In 2003 and 2004, a stem rust epidemic occurred in Kenya wheat growing regions rendering about half of the previously known Kenyan cultivars and Ug99 resistant global wheat materials susceptible especially those having *Sr*31 gene (Wanyera *et al*,. 2006). Race Ug99 (TTKSK) was responsible for it. In the subsequent years 2006 and 2007, two new *Pgt* races TTKST and TTTSK were detected to have virulence to genes *Sr*24 and *Sr*36, respectively.

Ever since, subsequent Ug99 variants have had virulence to other important stem rust resistance gene. Severe stem rust infection on recently released variety Robin having *SrTmp* was reported in 2014 (Newcomb *et al*,. 2016). It was a resistant gene effective to all known Ug99 races groups in Kenya. A similar epidemic occurred in 2013/2014, in the neighbouring country Ethiopia on a popular wheat variety (Digalu) that also had *SrTmp*. Results showed that this was not due to a new variant of Ug99 but another race (TKTTF) belonging to a different genetic lineage Clade IV with further subgroups of IV-A and IV-B compared to Ug99 and its variants (clade I) (Olivera *et al*,. 2015). These findings showed that the pathogen is evolving. To date, 13 race variants of Ug99 have been documented in wheat growing regions in 13 countries with 10 of them found in Kenya (Newcomb *et al*,.2016).

For Kenya wheat breeding programme to improve, surveillance of Ug99 and its race variants needs to be monitored. This can be achieved through detection of the available variants and know how closely they are related and how to deploy new resistant genes in the new cultivars in relation to the Ug99 variants identified. The mode of approach is use of molecular tools which is robust, timely and precise compared to race phenotypic data.

1.1 Statement of problem

Durable resistance has been anchored on stacking multiple *Sr* genes in the wheat varieties. It has been a strategy used by most wheat breeders with an aim of providing broad resistance to stem rust disease. The rust pathogen has been known to evolve into new physiological races rendering the present resistance varieties susceptible to the disease. The mechanisms bringing about these new races can either be facilitated through asexual mutations or, recombination (parasexual and sexual). In addition to this, the changes in the *Pgt* population is also brought about by air flow which transport potentially new genotypes spores from one region to another thus increasing the number of races from different clades and subclades. With this mode of dispersal, the variants recognized in the Ug99 lineage are spreading beyond Africa to other parts of the world. The *Pgt* population structure as well as the contributing factors of evolution of new physiological races in Kenya is not very clear. This has been attributed by the stem rust's mode of reproduction which is dynamic and complex in nature. The *Pgt s*pores responsible for infecting wheat may either be brought about by sexual cycle on an alternate host (*Berberies* sp) or asexual cycle from susceptible wheat varieties. These already existing spores on the susceptible wheat variety may evolve into new race via mutation. In most cases both alternate host and primary host do not coexist in one locality. The end result of either of these two cycles is emergence and re-emergence of new stem rust races which increases incidences, severity and the infection type of the disease on the already existing commercial cultivars. Understanding the driving force of the *Pgt* population structure as well as its genetic diversity could help in

strategizing breeding programmes for resistance and recommending specific genotypes for specific regions.

1.2 Broad objective

To contribute to improvement of Kenyan wheat production by examining population structure of stem rust pathogen 2015 in major wheat growing regions which will help inform Kenyan breeding program.

1.2.1 Specific objective

1. To determine the genetic structure of *Pgt* population in Kenya 2015.

2. To determine the driving force of the level of genetic variation of *Pgt* population in Kenya 2015 as to whether is brought about by asexual mutation, sexual recombination or migration forces.

1.3 Hypothesis

H_a: The *Pgt* population in Kenya 2015 is comprised of a single population with low genetic diversity.

H_a: Genetic variation of *Pgt* population in Kenya 2015 is primarily driven by asexual mutation and migration.

1.4 Justification of the study

Wheat (*Triticum aestivum*) is among the most important crops grown in Kenya coming second after maize in terms of consumption. It is among the cereal crops identified to contribute to vision 2030 agenda under the food security agenda. Ever since the identification of TTKSK (Ug99) 2004, TTKST (Ug99+*Sr*24 virulence) in 2006 and TTTSK (Ug99+*Sr*36 virulence) in 2007, this pathogen has been constantly evolving into new variants which render the current resistant wheat cultivars vulnerable to this pathogen (Singh *et al*,.2011b). Recently released Kenyan wheat variety Robin which had *Sr*Tmp gene broke down. So far, 13 variants which belong to the Ug99 race group have been reported in 13 countries in parts of East Africa and Middle East. With these new findings, scientists have now delved into the interest of wanting to understand the pathogen virulence as well as race structure (Patpour *et al*,.2015). In total, 10 races have been documented in Kenya making it the hot spot of the number of *Pgt* races since the year 2004 to 2015 (Singh *et al*,.2015). With this vast information, little is still known about distribution and driving force for the Ug99 variants especially in wheat growing regions in Kenya. Through the application of DNA-based tools, population genetics studies of *Pgt* have been highlighted shedding more information on the pathogen. This has been the case since the development of SSR markers specifically for stem rust pathogen (Szabo, 2007). In addition to this, *Pgt* SSR markers have also been used to differentiate the Ug99 race group from other *Pgt*

lineages (Jin *et al*,.2008). Data generated from the utilization of the SSR markers, lead to the development of polymerase chain reaction (PCR)-based diagnostic tool highly specific to Ug99 genetic lineage (Szabo, 2012).

Wheat production can be raised through characterization and quantification of stem rust pathogen and evaluating their population structure using molecular tools (Patpour *et al*,.2015). This is because the tools have the capability to measure directly the genetic diversity. With this knowledge gene pyramiding can be achieved thus selection of efficient resistance gene is achieved and increased (Singh *et al*,. 2015). In addition to this, it does supplement the already existing traditional race phenotyping studies on stem rust races.

CHAPTER TWO

LITERATURE REVIEW

2.0 Wheat stem rust disease and its importance

Since the beginning of human civilization thousands of years ago, wheat stem rust has been a problem. Research based on Chester, 1946, the Israelites related to the epidemics caused by cereal rusts and smut inflicted upon them as punishment for their sins. In addition to this, Romans had a festival known as Robigalia which was established to protect cereal crops through prayer and sacrifice to rust gods. With the emergence of new skills and knowledge, Aristole and Theophrastus associated the cereal rust epidemic with warm, wet weather. This means that this is an ancient disease given the fact that wheat was first domesticated around 10,000 years ago marking the transition from hunting and nomadic lifestyle to agrarian life.

Consumption of wheat products have risen due to increased urbanization, growth in incomes and growing preference for wheat products. In addition, more demand always comes from growth of animal feeds (Ellis *et al* ,. 2014). This means that demand for wheat products is high while the supply of the wheat to come up with the products is low.

In Kenya, wheat production has been there from the beginning of the 20th Century (Martens, 1978). It is mainly domesticated and used for commercial baking. The products

derived are ‘chapati’, ‘mandazi’, cakes, biscuits, bread and confectionary. Since the identification of pathotype *Pgt* stem rust, wheat crop has been hindered from reaching its maximum production (Roelfs *et al,*. 1992). This is because the infection disrupts the translocation of nutrients to the developing heads thus resulting in shrivelled grains which are of low milling quality. In severe infection, the stem of the wheat is prone to lodge. (Agrios, 2005).

In 2004, race TTKSK (Ug99, with virulence to *Sr31*) caused severe epidemic in most wheat growing regions in Kenya (Wanyera *et al,*. 2006). Since then, Kenya has witnessed drastic reductions in yield due to emergence of other new combinations of virulence in *Pgt* races (Ellis *et al,*. 2014). Majority of well-known Kenya varieties including ‘Eagle 10’ and ‘Robin’ have become susceptible to the evolved new Ug99 races thus derailing their production capability of 50,000 per ha in terms of estimated planted acreage (Singh *et al,*. 2011a). With the emergence of a new race which broke down variety ‘Robin’ having resistant gene known as *SrTmp* in 2014, it is evident that the pathogen is evolving and breaking down the available resistant varieties to this stem rust disease thus affecting wheat production (Patpour *et al,*. 2015). This pathogen has been of prime importance due to its geographical distribution and capacity to form new races which attack previously resistant varieties.

This gap in production poses another challenge of import price as there is challenge in foreign currency reserve and annual trade balance. It has mainly been attributed by changing policies, institutions and market arrangements thus lowering incentive in the wheat value chain. This gap can be reduced or filled through improved technology of improved resistant variety to the Ug99 race and its variants.

2.1 Epidemiology of stem rust disease.

Stem rust is recognized macroscopically by its characteristic small dusty reddish-brown pustules which are oval in shape found on both the leaves and stems (Agrios, 2005). The ideal environment for the development of the disease is moderate nights and warm days having the range temp of 16 ^{o}C - 27 ^{o}C (Roelfs, 1992). Sporulation takes place after 7 to 15 days once the infection sets in. The pustules on the stem are small lesions elongated vertically containing thousands of red-brown urediniospores which later change to black layer of teliospores as the crop matures. The infected crops are prone to lodging before harvest and harvested seeds from infected plant are usually shrivelled and light in weight (Roelfs *et al*,. 1992). Microscopically, these red spores are covered with fine spines.

Symptoms on the alternate host is usually characterized with special flask-like structures known as pycnia appearing on upper leaf surface of barberry plant (Roelfs,1985). They usually exude sticky like honeydew containing pycniospores. After 10 days later, the powdery aeciospores break through the lower leaf surface from the cup shaped structures. This is the macroscopic characteristic. On the other hand, the microscopically, warty surface is usually observed from the aeciospores.

2.2 Life cycle of stem rust *Puccinia graminis*

Pgt belongs to rust taxa which exhibit all five spore forms thus termed macrocyclic. It is a heteroecious rust fungi with susceptible wheat variety being the telial host and *Berberis* species acting as the aecial host. Infection by *Pgt* urediniospores which are diakaryotic (n+n) in nature and usually rust-coloured spores, are released by uredinia structure during warm months. They are spread via rain or wind thus repeatedly re-infect other susceptible wheat varieties (Roelfs, 1985). As the crop approaches end of growing season, a resting

spore stage termed as telia produces two-celled teliospores which is thick-walled. It is a dikaryotic cell with each cell having two nuclei. Karyogamy (fusing of nuclei to form diploid cells) occurs during early maturity of teliospore though not dispersed from the telial pustule. The teliospores remains dormant until early spring where they germinate to produce short tube-like mycelium in the alternate host. After karyogamy, meiosis occurs shortly which is thereafter suspended the teliospore dormancy period at diplonema of the first meiotic division (Boehm *et al*., 1992). Ultimately this process leads to production of four monokaryotic hyphal protrusion known as basidium. As the DNA replication and division continues there is formation of the basidium structure whereby the formed sterigma on each basidium forms new basidiospores expanding at the tip of sterigma. Sterigmata releases mature basidiospores and spread via wind to infect alternate host *Berberies* spp (Roelfs, 1992). The alternate host infection by the basidiospores is characterized by pycnium: small flask-shaped fruiting bodies that is formed in clusters on upper surface of barberry leaf which exude pycniospores that are contained in a solution "nectar" which are attractive to insects (Leonard and Szabo, 2005). These spores serve as male gametes while the other part of pycnia known as flexuous hyphae serves as female gametes. The fusion of these two gametes leads to formation of cup-shaped dikaryotic aecium below the pycnium finally rupturing lower epidermis of alternate host. Aeciospores which is diakaryotic produced by aceium is responsible for infecting the primary host which is susceptible wheat variety thereby completing the cycle.

2.3 Control of stem rust

Stem rust disease has been successfully controlled by finding and deploying new *Sr* genes against the Ug99 variants (Park *et al*,. 2011). The strategy used is stacking multiple

resistant genes alternatively known as gene pyramiding with an aim of having broad resistance against the disease. Currently, research work has been focused on understanding resistant present in breeding materials, the relationship between sources of different resistant (e.g target of the pathogen, linkage and broadness of resistance) and finally tools used in implementing the breeding strategies such as quantitative trait locus (QTL) mapping. The ultimate agenda is understanding host-pathogen system. Presently, there is no commercial varieties in Kenya with adequate resistance to the Ug99 and its variants. Application of fungicides as foliar has greatly played a role in the integrated management of the disease. Spraying the fungicides only reduces the disease build up but does not eliminate the disease. It is therefore considered as a short-term control of the disease provided the infection is not severe. The bulk of fungicides used in controlling wheat rusts are in the triazole class with a few in the strobilurine family (Wanyera *et al*,. 2016).

2.4 Phylogeny of stem rust *Puccinia graminis*

Rust fungi belong to genus *Puccinia*; order *Pucciniales*, phylum *Basidiomycota* and family *Pucciniacea* (Toome-Heller,2016). Studies done on rust fungi specifically on *P. graminis* has shown that it is monophyletic and a genetically variable complex species. (Zambino and Szabo, 1993). *P. graminis* has been connected to cereals as their commonly studied host however they can also infect forage and wild grasses (Leonard and Szabo, 2005). Most of the grass host of *P. graminis* belong to the group *Pooidea* and families *Triticeae, Aveneae* and *Poaceae*. It has a vast host range with more than 365 different grasses (Leonard and Szabo 2005). The range include telial hosts from some 77 genera of *Poaceae* and more than 70 species of *Berberis sp* and *Mahonia sp* as aecial hosts (Roelfs,1985). *Puccinia graminis* is divided into subdivisions based on host specificity (*formae specials*)

or morphological differences (subspecies, varieties, forms). *Formae speciales* (special forms: f spp) is usually named according to the associated telial host. In *P. graminis* those recognized include *P. graminis*. f. sp *avenae* on oats (*Avena sativa*), *P. graminis*. f. sp. *lolii* on ryegrass, (*Lolium*), *P. graminis*. f. sp. *secalis* (*Pgs* on cereal rye *Secale cereal*) and *P. graminis*. f. sp. *tritici* (*Pgt* on wheat *Triticum aestivium*) (Leonard and Szabo, 2005).

Urban, 1967 highlighted two subspecies *graminicola* and *graminis*, whereby the former is found on non-cereal grasses while the latter is found on cereal crops.

2.4.1 Definition of races/race nomenclacture of *Puccinia graminis* f. sp. *tritici* (*Pgt*)

Races of *Pgt* are determined by phenotypic avirulence/virulence testing on a standard set of wheat differentials. The race nomenclature of *Pgt* races has evolved from chronological numbering system by Stakman (Stakman and Levine 1922) into current letter system (Roefls and Martens 1988). The system is based on the infection types expressed by different isolates of *Pgt* produced on defined wheat line. Infection types are classified based on a scale of 0 to 4 with; 0, no visible infection, ";", flecking, 1, small uredinia often surrounded by necrosis, 2, small to medium size uredinia surrounded by necrosis or chlorosis, 3, medium uredia associated with chlorosis and 4, large uredinia without chlorosis or necrosis (Roefls and Martens 1988). Infections types designated as 0, ";",1, or 2 are classified as resistant reaction and 3 or 4 to be susceptible reaction.

A standard system for race identification was published by (Roefls and Martens 1988) which is currently known as North American stem rust nomenclature. It is a system based on utilizing single gene wheat lines in sets of four and assigning letter codes based on resistance and susceptibility to rust. Improvement has been made from the previous 12 lines

(three-letter code) to 16 wheat lines which has four-letter code (Roefls and Martens 1988). More recently it has been improved to 20 lines (five-letter code) (Jin *et al* ,.2008).

2.4.2 Recent races of *Pgt* in Kenya.

In the past stem rust control has been effective through use of the stem rust resistant gene cultivars (Ellis *et al*.,2014). This strategy has been derailed since stem rust races TTKSK, TTKST and TTTSK caused severe epidemic in Kenya the years 2004, 2006 and 2007 respectively (Wanyera *at al*,. 2006, Jin *et al*,. 2008). To date, among the 13 known race variants of Ug99 which belong to Clade I, 10 have been found in Kenya (Table 2.0). From the d-samples analysed from Kenya as from 2015 to 2017, it was noted that there is entry of new race TTTTF and TKTTF both belonging to Clade IV-B, different from the Clade I race (Pablo D. Olivera and Yue Jin, 2018). (Table 2.0)

Table 2.0 Races in Kenya as from 2004 to 2017 with inclusion of both Ug99 and Non Ug99 race groups

Clade	Races	Year of existence
Clade I (Ug99 race group)	TTKTT, PTKST, TTKTK, PTKSK, TTKSK, PTKTK	2004 to 2015
	TTHST, TTKST, PTKSK, TTTSK	
Clade IV-B (Non Ug99)	TTTTF	2015
	TKTTF	2017

The neighbouring countries sharing races present in Kenya are Uganda, Tanzania and Ethiopia. Uganda has races TTKSF, TTKST, TTTSK, TTKTK and TTKSK. In the case of Tanzania races found there are TTKSK, TTTSK and TTKST. Lastly Ethiopia has races TTKSK, PTKST, TTTSK, PTKSK and recently TKTTF (Singh *et al*,. 2015, Olivera *et al*,. 2015). In addition to this, the Ug99 race group has spread to other more 9 countries in

Africa and Middle East with emphasis on South Africa, Eritrea, Yemen, Iran, Mozambique and Eygpt. Ug99 races TTKSK, TTKSK, TTTSK, PTKSK, PTKST, TTKTT, TTKTK, TTHSK, PTKTK and TTHST have shown broad virulence to *Sr*31 while race TTKSF, TTKSP and TTKSF+ is avirulent to *Sr*31. Races TTKST, TTKSP, PTKST, TTKTT and TTHST is virulent to *Sr*24 while PTKSK, PTKST and PTKTK is avirulent to *Sr*24. Lastly, race TTKTT, TTKTK and PTKTK has shown virulence to *Sr*Tmp. (Jin *et al*, 2007, Singh *et al*,.2015). This is a brief profile of the key *Sr* genes deployed in wheat varieties to control Ug99 race group.

2.5 *Puccinia graminis* f. sp. *tritici* genome

The average number of predicted genes of rust fungal genomes ranges from 15,000 to 20,000. The overall assembly size of haploid genome of *Pgt* is 88.6Mb having transposable elements (TEs) occupying approximately 44% of the assembly with the prediction set of 17,773 protein-coding genes (Duplessis *et al*,. 2011). The current update on this data indicates that the predicted protein-coding genes has been revised to 15,800 accounting to approximately 36.5% of the genome (Cuomo et al., 2017). The size of this proteomes is similar to symbiotic basidiomycete *Laccaria bicolor* but relatively larger than non-obligate pathogenic biotrophs: corn smut fungi *Ustilago maydis* and *Sporisorium reilianum* which encode for approximately 6,500 proteins (Martin *et al*., 2008, Schirawski *et al*,. 2010).

2.6 Overview of the Molecular markers

We can trace back the science of plant genetics during the Mendel's classical studies on garden peas. Since then it has formed the basis ground for scientist to constantly identify, sort and map for single gene markers in many higher plant's species. It was identified that

the Mendelian factors do control inheritance which currently is known as genes and arranged in linear fashion on chromosomes (Bateson W, 1909). Molecular markers have been designed to show the genetic differences on a more detailed level with the exception of inferences from environmental factors. These are the chromosomes segments which neither encode any traits nor affected by environment but only inherited in Mendelian fashion (James and John, 2011).

In general, the markers have to be polymorphic (existing in different forms) with an aim that the chromosomes carrying mutant gene is easily distinguished from chromosome having normal gene and form of marker it carries. The different forms of the markers can easily be detected at three levels. These include: phenotypic level, difference in proteins or difference in the nucleotides sequence of DNA. In summary they are morphological markers, biochemical markers and lastly molecular markers.

Globalization trend in agriculture has recently created various aspects of plant protection. This has brought the sense of in-depth studies of host-parasite interactions, disease resistance, and pathogens population structure (Atallah *et al*., 2010). For these fields to be extensively understood there has been a need for development of methods capable to detect and identify pathogens in question present in plants in a fast, accurate and sensitive manner (Monda and Shanmugam 2013). Visual identification of plant pathogens is rapid and cost-effective method of disease diagnosis. The only limitation is that, it requires experienced personnel and limited only to disease affecting aerial parts of plant. The use of protein/nucleic acid-based detection methods is robust and repeatable. They can overcome the problems associated with microscopical detection of plant pathogens. This is because they are useable at any developmental stage of plant due to the fact that all living cells containing entire set of genomes is not affected by environment (Atallah *et al*., 2010).

Introduction of the molecular markers' techniques in the field of plant pathology, has greatly improved the ability of detection of plant pathogens and upgrade of understanding of their epidemiology and ecology (Grünwald *et al*., 2003).

In the last 30 years these markers have been extensively been used with a lot of upgrading. Their usage and classification have been based on generated sequence information from generic DNA fingerprinting and sequence specificity to alleles in the plant genome (Mago *et al*., 2011). In the early 1980s first generation DNA marker were developed. These included restriction fragment length polymorphism (RFLP). Its basis of functionality was based on restriction digestion and hybridization with probe. It also used large amount of high DNA purity to run. A few years later random amplified polymorphic DNA (RAPD) markers come into existence. It's mode of functionality was random primers for PCR amplification (Kumar *et al*., 2010). The challenge identified with this marker tool was that it was unrepeatable and unreliable.

In the year 1990 Amplified fragment length polymorphism (AFLP) was introduce to solve the deficiency created by the RAPD as it improved on its reliability (Skrede *et al*., 2012). It highlighted on the detection of genomic restriction fragment by PCR amplification. More advanced marker tools were later developed in the early years of 2000 that focussed on the microarray-based DNA fingerprinting methods (Mondal and Shanmugam 2013). These included simple sequence repeats (SSR), sequence-tagged site (STS) and allele-specific PCR (AS-PCR) among others. Their applicability was focussed on sequencing specificity to loci in genomes (Gupta *et al.*, 2000). The fact that they would tolerate high levels of impurities in crude DNA extracts makes them more advantageous in terms of being robust.

In addition, efficient breeding requires the use of robust molecular markers as it accelerate the genetic wheat improvement. This is to help in creating genetic variation of beneficial

traits (Zhong *et al*., 2009). The thought has been attributed by the growing human population who exert pressure on the available arable land. Achieved crops containing the artificially inserted genes will be of great economic benefit to wheat farming communities in terms of resistance to pathogens and yield improvement (Blaszezyk *et al*., 2004).

The criteria used in choosing these markers include polymorphism and its distribution in crop genome, cost effective genotyping procedure and the high tolerance of impurity in DNA samples from extracted methods. Finally, the amenability to automation efficiently of large crop sample is also a consideration factor (Yang *et al*., 2015).

2.7 Use of molecular markers in plants and pathogens genomes analysis.

2.7.1 Restriction Fragment Length Polymorphism (RFLP)

This molecular tool was originally developed for mapping human genes. It is one of the first-generation molecular markers to be designed and used. However, there has also been molecular markers generated for plants, bacteria and nematodes (Mondal and Shanmugam 2013). RFLP markers have been designated into two types: one associated with gene of known function while the second one associated with anonymous gene segments of one sort or another. The principle behind this technique is that the DNA is digested within two restriction endonucleases. The DNA sample is broken up and digested by restriction enzyme. The DNA fragments is later separated by length on agarose gel electrophoresis and determining the number of fragments and relative sizes. The detected RFLP is later visualized using X-ray film in an autoradiography. In summary the whole RFLP process requires probe labelling, DNA fragmentation, electrophoresis, blotting, hybridization, washing and autoradiography. In the past this technique has also been used in detection and characterization of fungi (Woo *et al*.,1996). Advantages associated with this technique is

that if location of a particular disease gene is being sought for anchored on specific chromosome, the family DNA sample with the disease would be analysed. The similar patterns of inheritance in the RFLP alleles would be looked for. Some of the disadvantages associated with RFLP are that they are expensive, laborious and time-consuming process thus making it obsolete. In an addition to that they are also amenable to automation. They also require qualified personal.

2.7.2 Random Amplification Polymorphic DNA (RAPD)

This is one of the simplest techniques as short PCR primers (approximately 10 bases) are usually randomly selected and amplified throughout genome. They are single primers of arbitrary nucleotide sequence. Application of this technique is development of genetic maps, genetic markers and phylogeny information (Monda and Shanmugam 2013). The visualization is done on agarose gels stained with ethidium bromide. In Saudi Arabia it has been used to highlight genetic diversity of *Capparis deciduas* (Abdel-Mawgoud *et al*., 2010). (Kumar *et al*., 2010) have used this tool to identify the rice blast pathotypes and the resistance genes that can be convened in rice. RAPD have the advantage to simultaneously detect polymorphic loci in various regions of a genome. They are quick to detect, requires small amount of template DNA and no requirement of DNA sequence information. In addition to that they can be viewed using fluorescence. The noticeable disadvantage is that they are anonymous and their level of their reproducibility is very low due to non-specific binding of short, random primers.

2.7.3 Amplified Fragment Length Polymorphism (AFLP)

AFLP are anonymous having high level of reproducibility and sensitivity being high due to their longer +1 and +3 selective primers. The principle behind its amplification is based on selected restricted fragments of a total digested genomic DNA. The amplified products are later separated by electrophoresis once labelled on a range of 60 to 500 base pairs (Mondal and Shanmugam 2013). This nature makes them applicable to all species. They have been designed to synthesize first arbitrary primers then ligate them to target DNA fragments. AFLPs are treated as dominant markers. Presence of large numbers of fragments gives an estimate of variation. When based across the entire genome it highlights the view of genetic variation of the studied organism. The length of the fragmented DNA polymorphism has been successful been used to differentiate several bacterial species (Mondal and Shanmugam 2013). They are also mostly used in determining genetic diversity and phylogenetic relationship between the closely related genotypes (Yang *et al*., 2015). The presence of discriminatory nucleotides at 3' end of each primers also makes them unique. However, they have disadvantage of being lengthy and laborious in detection method which are neither amenable to automation. Their development and operational cost is also relatively high.

2.7.4 Simple Sequence Repeats (SSR)

SSR are molecular makers which are more advanced than the above discussed molecular tools. They are short sequence of DNA repeats in a specific number of times. The repeats are usually of short nucleotide sequences equal to or less than six bases in length varying in numbers (Gupta *et al*., 2000). This repetition varies between plants within species thus indicating the presence or absence of an allele which is mainly attributed by their high polymorphism arising from mutation rate.

This tool has been used in the analysis of stem rust pathogen with the help of 20 developed polymorphic and highly specific dinucleotide SSR markers (Szabo, 2007). It is a wheat pathogen which has elicited a lot of interest due to the emergence of new races that are closely related (Park *et al*., 2011). The results yielded by these primers stated that expected levels of heterozygosity ranged from 0.000 to 0.960. This proved that they are specific to the pathogen thus molecular markers of choice. In addition to that, they have been used to state the pathotypes relatedness with preciseness (Visser *et al*., 2011).

They have been used widely due to their easy use by simple PCR which is followed by denaturing gel electrophoresis for the alleles per locus as they are evenly distributed in the genome (Mondal and Shanmugam 2013). The large sizes generated do provide high degree of information on number of alleles per locus. This is because they are no longer anonymous but highly polymorphic and reproducible and finally amenable to automation. Lastly little DNA is required making it easy and cheap to run. The coupled advantages make results to have expected heterozygosity frequency of being greater than 0.7 thus allowing for precise discrimination even of the closely related individuals. The only challenge is that they are costly in terms of financial (Hammer *et al*., 2000).

2.7.5 Single Nucleotide Poylmorphism (SNP)

SNP has become a favourable molecular tool owing to their wide distribution in genomes and compatibility with high, multiplex detection systems (Akhunov *et al*., 2009). This means that there is a single base change in a DNA sequence with the usual alternative of two possible nucleotides at a given position thus causes different alleles. This nature makes them bi-allelic co-dominant markers.

Recently they have been replacing the SSR markers in plants due to them being more abundant, stable, amenable to automation and increasingly cost effective. In recent times, SNPs have been used in molecular comparison especially for evolutionary and ecological (Choi *et al*.,2001, Kawuki *et al*.,2009). The technology has also come in handy for race typing the organisms having high mutation rates especially RNA viruses where mutations are presumed to be caused primarily by an error-prone RNA dependent RNA polymerase (Choi *et al*.,2001). In the study of plant pathogens, bacteria and fungi are presumed to be the most predominant. With time their genetic analysis has posed a challenge due to their genome size, multiple life stages, diploid genomes, presence of plasmids and finally climate change which encourages rapid mutations (Agrios *et al*., 2005). Plant pathologists can adapt this tool in disease diagnosis through the application of what is known about organism's life cycle, genomic features and mutation rates to the genetic marker section and reaction conditions. In rice, the SNP discovery on genome wide basis has been used based on either genomic scale re-sequencing approaches or Sanger sequence-based strategies. In oat genome a high-throughput SNP discovery has been developed based on resolution melting and 454 genome sequencing technology (Yang *et al*., 2015).

The SNP genotyping technique is usually based on generation of allele specific molecular reaction products, their separation and detection of allele specific products for their identification. They are faster than SSR test allowing for more tests to be done in a short period of time. It has made it possible to create saturated maps which enable easy genome-wide tracking, fine mapping of target regions, and rapid association of markers with a trait and finally expedition of cloning genes or quantitative trait loci of interest (Monda and Shanmugam 2013).

The currently available markers are being use for genotyping basing on comparison of locus-specific sequences generated from different chromosomes. International centre for maize and wheat improvement (CIMMYT) researchers are now striving to develop and select the wheat germplasm that have broad spectrum resistance to the new races identified (Singh *et al*., 2011). This has been intended to be achieved through stacking several new resistance genes into each new adapted variety (Mago *et al*., 2011).

Despite there being extensive Multi-locus DNA fingerprinting methods (RAPDs, AFLPs, SAMs and STMs), there application to rust pathogens has been limited due to dikaryotic nature of rust fungi. In spite of this being the case, SSR markers have withstood the challenge and extensively used in understanding genetic diversity of *Puccinia* spp (Szabo 2007; Visser *et al*., 2009). This marker has been used to demonstrate distinctiveness of Ug99 and its associate races found in different parts of Africa (Jin *et al*., 2008; Jin *et al*., 2009; Visser *et al*., 2010).

2.8 Review on *Pgt* genotyping work

In the past, Random amplified polymorphic (RAPD) method was modified and used in the genetic analysis of *Pgt*. It emphasised on utilization of primers with G+C content of 80 to 100%. The results yielded showed that these primers resulted twice the number of amplified products compared to standard 60-70% G+C primers. High GC primers were generated with respect to segregation and similarity based on genetics (Kubelik and Szabo,1995). In addition to this, these markers have been used in the comparison of twenty and nineteen isolates of *Pgt* from South America and Europe respectively. This marker showed a level of similarity between the two continents isolates with average similarity coefficient of 0.70 (McCallum *et al*,.1999). Amplified fragment length polymorphism

(AFLP) markers showed the genetic linkage of avirulence loci in *Puccinia graminis* that control avirulence and virulence on different *Sr* genes (Zambino, *et al*,.2000).

Keiper *et al*,. 2003 in the past has also utilized a combination of high multiplex DNA fingerprinting techniques to assess the genetic relationship among isolates in Australia focusing on five cereal rust pathogens, *Pgt* being one of them. DNA fingerprinting techniques utilized included amplified fragment length polymorphisms (AFLP), selectively amplified microsatellites (SAM) and sequence-specific amplification polymorphisms (S-SAP). SAMs markers were the most informative in comparison to the other two markers. This is because it had the highest marker information content based on number of species-specific fragments, percentage of polymorphic loci, marker index and polymorphic fragments. It has been identified as a marker having potential for the development of locus-specific microsatellites. In the subsequent years 2006, 110 microsatellites for *Pgt* were developed through Sequence-tagged microsatellite profiling. It was used to understand the genetic diversity of *P. graminis* f. sp. *avenae* (causal agent of oat stem rust) isolates in Australian populations (Keiper *et al*,.2006). The results revealed presence of subpopulations in the pathogen population. This marker led to the development of twenty-four dinucleotide simple sequence repeat (microsatellite) markers for *P. graminis* which were evaluated using 25 isolates of *Pgt.* From the preliminary results, this *Pgt*SSR primer pairs indicated to be specific to *P. graminis* with respect to other cereal rust fungi (P. *coronata*, *P. striiformis* and *P. triticina*). Polymorphic loci were identified with a range of two to eleven alleles on allelic diversity (Szabo, 2007). In South Africa, *Pgt*SSR markers has been utilized to highlight *Pgt* genetic structure as well as identify new races of stem rust of wheat in the Ug99 lineage (Visser et al., 2009: Visser *et al*., 2011). Finding showed division of the population in two distinct groups while in the latter research detection of

TTKSP and PTKST race forming genetic relationship within Ug99 lineage. Ethiopia *Pgt* genetic diversity was shown to be high with genetic differentiation among population missing based on geographical separation as per the reflection of low coefficient (≤0.046) (Admassu *et al*., 2010). Stoxen 2012, screened ninety-nine trinuclotide SSR markers for the *Pgt* race dynamics genotypic analysis in United States. He later settled on twenty of the markers for his analysis basing on polymorphism between isolates among other factors. In the recent time, two illumina custom SNP arrays *Pgt* SNP 1.5k and *Pgt* SNP 3.0k chips have been used to characterize the *Pgt* samples. (Olivera *et al*,. 2015) genotypically characterized race TKTTF from Ethiopia using *Pgt*SNP 1.5kb while the Kenyan isolates samples from 2013 and 2014 were characterized using *Pgt* SNP3.0kb (Newcomb *et al*,. 2016). These results, did confirm that all Kenyan isolates were clustered in clade I: reference of Ug99 race group isolates. Olivera *et al*., 2017 analysed 40 isolates of *Pgt* from Germany wheat fields 2013 using custom SNP array whereby the phylogenetic analysis showed two distinct lineages : clade IV and clade V.

From the past results shown by these molecular diagnostics they have proved to be rapid and robust clearly demarcating the population structure of stem rust in host research countries.

CHAPTER THREE

MATERIALS AND METHODS

3.0 General description of major wheat growing areas of production.

Kenya has ten predominant wheat varieties which are grown in major wheat growing regions. They have different attributes that favours their adaptation to these environments. Below are general highlights of some of the notable varieties with special attributes that Kenya wheat farmers look for in different agro- climatic zones (Table 3.0).

Table 3.0: **Kenya wheat varieties, grown regions and their attributes**

Production region	Variety	Altitude (masl)	Yield Potential (Tons/Ha)	Special Attribute
Njoro, Timau, Molo soils	Kenya Wren	1800-2400	8.5	Tolerant to acidic
Mau-Narok Njoro, Timau, Molo Narok, Eldoret	Robin	1800-2700	8.1	Widely Adapted
Njoro, Timau, Narok baking	Kenya Sunbird	1800-2100	6.5	Good for industrial
Mau Narok, Molo, Eldoret				
Njoro, Rongai, Lower Narok, Kinamba, Naivasha	Eagle 10	2100-2400	6.5	Early maturity

Njoro, Timau, Molo, Mau Narok, Upper Narok, Eldoret	Kenya Korongo	2100-2400	8.5	Excellent baking qualities
Njoro, Mau, Mau Narok, Lower Narok, Molo, Kinamba	Kenya Kingbird	1800-2400	6.0	Early maturity
Njoro, Timau, Mau Narok, Narok, Molo, Eldoret	Kenya Hawk12	2100-2400	8.0	Pre-harvest sprouting tolerant
Njoro,Timau, Mau Narok, Upper Narok, Molo, Eldoret	NjoroBW2	2100-2400	8.0	High yielding
Njoro, Timau, Narok, Upper Narok, Molo, Eldoret	Kwale	2100-2400	7.5	High biomass
Njoro, Rongai, Lower Narok, Naivasha	Duma	1800-2100	6.0	Drought tolerant

Source: Kenya Wheat Production Handbook (2016)

GPS Coordinates

3.1 Collection of wheat stem rust samples

Initial plan was to randomly select four farms from each wheat agro-ecological zones and collect approximately 50 samples of wheat infected with *Pgt.* This was not possible as most farms had been sprayed with fungicides to control the disease. A total of 536 samples were collected from the four major wheat growing regions; North Rift (108 samples), South Rift (156 samples), Central Rift (171 samples), and Mount Kenya (101 samples). Detailed information on the farms visited is given in (TABLE 3.1).

Table 3.1 Location of 10 Kenyan wheat field collection inclusive of wheat variety grown in each field.

Region	Field	Wheat variety grown	Elevation (MASL)	Latitude	Longitude
NORTH RIFT	MEREWET	Robin (*Sr2, SrTmp*)	2092m	N: 00. 41161	E: 035.18385
	RAY FARM (SERGOIT)	Robin (Sr2, *SrTmp*)	2277m	N: 00. 45116	E:035. 28318
SOUTH RIFT	NAROK OLOPITO	K.Hawk 12 (*Sr2, SrTmp*)	2106m	S: 01 00292	E:035 53241
	MBOKISHI	Robin (Sr2, *SrTmp*)	2005m	S: 00 59474	E:035 42294
	NGORENGORE	K.Hawk.12 (SrSrTmp)	1983m	S: 01.01596	E:035.3706
CENTRAL RIFT	KWA NDUSU RONGAI	Robin (Sr2, SrTmp)	2058m	S: 00. 15520	S: 035. 53418
	MENENGAI	Robin (Sr2, *Sr Tmp*)	1969m	S: 00. 11428	S: 036. 00416
	PIAVE	Robin (Sr2, *Sr Tmp*)	2137m	S: 00. 21238	E: 035. 58311
MT KENYA	NGUSHISHI-MARITATI	Robin (SrTmp)	2456m	N: 00.05090	E:037.18300
	MAGUTU-LAIKIPIA	Robin (*Sr2, SrTmp*)	2102m	N: 00.05332	E:037.11279

[a] MASL, Meters above sea level.

Depending on the regions sampled, approximate distance between sampled farms was 5 kilometres. A GPS (Global positioning Satellite) device was used to determine the coordinates of each sampled field. The fields were crisscrossed in order to identify infected stem with rust from the wheat plant. Stem sheath having single uredinial-pustule in the sampled farms were cut into small pieces of approximately 5 to 10cm in length then placed in an office envelope. The scissors used was disinfected using 70% ethanol before cutting new samples (infected wheat stem) as well as moving from one field to another. Wheat variety grown was recorded. The collected samples were stored at 4 ^{o} C until DNA extraction.

NM

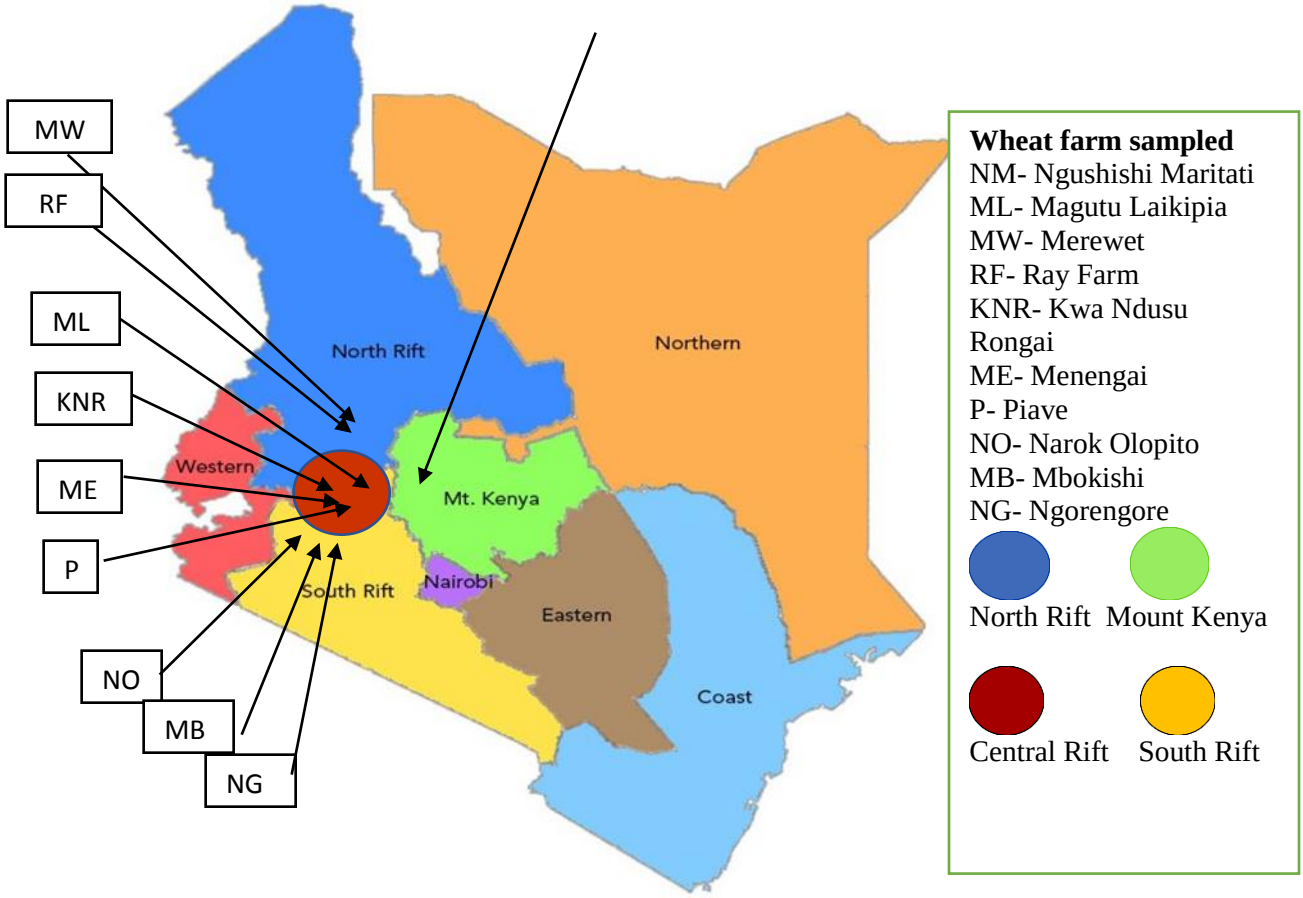


Fig 1. Map showing major wheat growing regions in Kenya and location of sampled sites (Source: Google Maps, 2019)

3.2 Molecular characterization of stem rust races and its diversity

3.2.1 DNA isolation

DNA was extracted from the single uredinial pustules using modified CTAB method by (Doyle and Doyle, 1990). Modification involved addition of 20g of Sodium Dodecyl Sulphate (1M Tris-Hcl pH8.0, 0.5M EDTA, 5M Nacl) (SDS) to 100mL of the CTAB extraction buffer (20% SDS). Prior to grinding the sample, the modified CTAB extraction buffer (CTAB, PVP,1M Tris-Hcl pH 8.0, 0.5M EDTA, 5M NaCl) was pre-heated at 65^0C in water bath for 30 min.

0.5-1g of sheath samples containing the single uredinial pustule was ground in a mortar having 500µl of extraction buffer with a pestle.100µl extraction buffer was added on pestle to rinse off rust that might have attached to it as well as further break down the stem sheath cell wall. The extract was transferred into 1.5ml microfuge tube then incubated for one hour at 65^0C in a water bath. The incubated samples were removed from the water bath, cooled at room temperature for 30 mins before adding 600µl phenol; chloroform; isoamyl alcohol (25:24:1) into the microfuge tubes. They were shaken until milky then centrifuged at 18,928 x g for 10 minutes. The supernatant 600µl was transferred into a new microfuge tube, mixed and then centrifuged at 18,928 x g for 5 minutes. The upper aqueous phase 450µl was transferred to new microfuge tube. To precipitate the nucleic acids, equal amount of chilled isopropanol was added, tubes carefully inverted and incubated for 15min on ice. After incubation, the tubes were centrifuged for 15 minutes at 18,928 x g at room temperature and the supernatant was discarded. To each of the tubes, chilled 500µl of ethanol (70%) was added, then inverted for 5 minutes centrifuged for 15 minutes at 18,928 x g at room before decanting off the ethanol. The remaining pellet was air dried by inverting tubes on paper towel for 30 min and then dissolved in 50µl Tris EDTA (TE) buffer (10mM Tris HCl, 1 mM EDTA [pH 8]) for DNA purification and storage. Finally, 2µl of DNAase-free RNase A (20mg/ml) was added and the samples incubated at 37 ^{o}C for 30 minutes before being stored at 4 ^{o}C.

3.2.2 DNA Quality.

NanoDrop spectrophotometry was done on the samples to assess the purity and quantity of DNA. This entailed the nanogram/microliter (ng/ul) i.e. how much DNA and RNA were present in the sample and 260/280 ratio respectively. The criteria for the choice of samples

used in running the PCR reactions was based on samples having the 260/280 ratio of more than ~1.8 even if the DNA quantity nanogram/microliter (ng/ul) was low.

3.3 Markers used for genetic analysis .

DNA samples were genotyped using *Pgt*SSR markers. Two standard reference isolates were used: 04KEN156/4, race TTKSK, clade I; 13ETH43-1, race TKTTF, clade IV-B. These standard isolates were collected from Kenya and Ethiopia respectively. For the purpose of this study, 20 *Pgt* SSR markers were used; Pgest142, Pgest227, Pgest293, PgtGAA8, PgtCAA53, Pgestrre024, PgtCAA80, Pgest109, PgtCAA49, Pgest318, PgtCAT4_2, Pgest059, Pgest173, PgtCAA93, PgtCAA39, Pgest341, Pgest021, Pgest098, Pgest325, Pgest353 (Stoxen S, 2012). (Table 3.2)

Table 3.2 Primer sequences and characteristics of the 20 SSR makers used in the genetic diversity study of Kenya *Pgt* samples 2015.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Super contig[a]	Size range (bp)	No. of repeats	Motif	Reference
PgtCAA39	CGTCGTCCCTCCATAGTCTTA	CTCTCAAGCACCCTCAACATC	2.11	197-245	13-29	CAA	Stoxen S, 2012
PgtCAA49	TCGTCTGATCGTGAGAAACG	GACGATTGCTGAGGATTGCT	2.21	131-152	13-20	CAA	Stoxen S, 2012
PgtCAA53	AGGCTCAACACCACCCATAC	AGGAGGAGGTGAAGGGGATA	2.14	215-239	8-16	CAA	Jin et al. 2009
PgtCAA80	GCCTCCAGACGAATGGTTTA	TTGGTGATGATGATGGTTGG	2.3	239-248	10-13	CAA	Stoxen S, 2012
PgtCAA93	CGCCTGTGATGGTTGTATTG	CACTCTCGCCAAACCTCATT	2.43	155-182	5-13	CAA	Jin et al. 2009
PgtCAA98	ATTCGGATGGTCCGTTACTG	CCATCCCACTCAAATCATCC	2.10	153-180	8-17	CAA	Jin et al. 2009
PgtCAT4.2	CCGTGTCGATCCCAATAATC	AGCAAGGTGAGAATCGGAAA	2.16	123-144	8-15	CAT	Stoxen S, 2012
PgtGAA8.1	TGTCTGCCTGTCTGTCGAAC	GGATGATCGGTCAGTTGGTT	2.33	188-209	5-12	GAA	Stoxen S, 2012
Pgest21	CCGAATGCAGATTACCCTTG	GTTTGCCTGATGATGGATGA	2.2	246-255	6-9	AAG	Stoxen S, 2012
Pgest24	TCATCGACCAAGAGCATCAG	TTCGGGAGTGAGTCTCTGCT	2.3	157-160	7-8	CAT	Stoxen S, 2012
Pgest59	AGGTTGATGATGAGGATGC	ATTATGCGGGACAAATCGAG	2.5	215-221	5-7	ACG	Stoxen S, 2012
Pgest109	CCATCCGATCATTTCTTCGT	CCGACCTTCTCTTGCTTCT	2.13	170-176	5-7	CCT	Zhong et al. 2009
Pgest142	GATGGTGAAGTCCGGTATGG	CCACCAACAAACCAACAAGA	2.17	198-213	9-14	TTG	Stoxen S, 2012
Pgest173	TCCATTGAGTTCCATCGTGA	TCCCTTGACCATTCTCAACG	2.22	181-190	5-8	ATC	Zhong et al. 2009
Pgest227	CACACGTCTCGAGGAACAGA	CTCGTGGGATGAAGTCCATT	2.33	193-214	8-15	AAG	Zhong et al. 2009
Pgest293	GAACCTTGGCCTGAGTGCTA	GCAGCCTACAGCAAGAATCC	2.54	240-246	4-6	GGT	Zhong et al. 2009
Pgest318	GATGTCGGTCTTGGTCCACT	ACAGACACTCCCGAGCTCAT	2.64	256-274	5-11	ATG	Stoxen S, 2012
Pgest325	TTGGGTGAGTCAGAGTTTGAGA	CCCACCCACTCTCAGTCAAT	2.70	171-180	4-8	AAG	Stoxen S, 2012
Pgest341	GGCCTTGGTACCCAATTTCT	GATGTCGCACTCGGTTTCTT	2.83	214-217	4-5	TGG	Zhong et al. 2009
Pgest353	ACGTCTTGGGTTTCTGTGGA	TCGAATCCCAAGGAACAGAG	2.92	235-241	4-6	AGG	Zhong et al. 2009

SSR marker location information is from Stoxen S, 2012 based on the *Pgt* assembly (Duplessis et al, 2011).

PCRs amplification was carried out in a MJ Research PTC-200 Gradient Thermal Cycler (Marshall Scientific, Hampton USA) thermocycler. PCR reaction mixes contained the following: 1.0 µl of 0.05% Casein, 2.0 µl 5 X Phusion HF buffer with 7.5 mM $MgCl_2$ (New England BioLabs, Ipswitch MA), 0.2 µl 10 mM dNTPs (Roche, Indianapolis IN), 0.25 µl SSR_F' (10µM) and SSR_R' (10 µM) labelled primers, 5.25 µl ddH_2O, 0.05 µl Phusion Taq Polymerase and 1.0 µl DNA. PCR profile for all markers was by Touchdown (TD) method with 10 cycles of 3min at 95 ^{0} C,30 sec at 94 ^{0} C, 1 min at 63 ^{0} C, 30 sec at 72 ^{0} C; and then 30 cycles with 30 sec at 94 ^{0} C,1 min at 58 ^{0} C, 30 sec at 72 ^{0} C and a final extension at 72 ^{0} C for 20 min and then 10 ^{0} C indefinite. The amplified PCR products were transferred to a 96 plate well in the dilution ratio of 1:9 whereby 1 represented FAM dye or HEX dye used as ladder and 9 the amplified DNA from samples to make a total dilution of ten microliter. Among the SSR *Pgt* markers, 10 were FAM dye exhibiting 'blue colour' while the other remaining 10 were HEX dye exhibiting the 'green colour'. This was how I would differentiate one primer from another when reading the SSR peaks of the amplified products. The amplified multiplex PCR products in the 96 well plates were taken for fragment analysis using the ABI 3730xl capillary electrophoresis platform (the same used in Sanger sequencing services) at University of Minnesota Genomics Center. The results which exhibited the peaks (Chromatograms) were read by Geneious Microsatellite Plugin 1.4 for Geneious R7 software. Before calling of any alleles, the ladder was checked to be called correctly. The dye showing the ladder was red in colour. Secondly the microsatellite loci information was set. In the case of this data, 6-FAM dye was used having the characteristic of two expected peaks with four repeat units. The start range was 160 while stop range 450. One peak from the reading showed that the sample was homozygous while two peaks signified that the sample was heterozygous. In most cases the

homozygous peak size doubled the heterozygous peaks size combined. Stutter peaks were minor peaks that were repeated 1-4 units away from the actual peaks. This is how the data was read based on the called peaks shown. The amplification of PCR products was replicated into two so as to confirm the peak reading as well as cover up of the failed reactions. The failed reactions were also repeated twice so as to get exact chromatograms readings.

Before data analysis, binning of the data was done. It was guided by the rule that SSR alleles should always be separated by 3 base pairs and the SSR data with decimal values for base pairs binned or changed to integer values. In the case of a monomorphic allele, where only one allele was seen, the same integer was coded twice. At times, the nature of analysis is that there is always some uncertainty in the calls (electropherograms) and therefore a call of 148 and a call of 149 are actually represented by the same allele, which we can score 148. The main aim is to have consistency for all DNA samples analysed by the marker. For example,

139	142	145	148	151	154	157	160

In the case of the correction for binning of observed allele size with the actual allele size based on DNA sequence data, spreadsheet was created having the electropherograms calls/readings. Calculations was done by subtracting 3 from every cell to the left and adding 3, to every cell to the right, continuing out from a central cell. The central cell was changed when we decide to re-evaluate bin size based on previous PCR reactions conducted using the same marker. The goal of this part of the re-binning analysis was that no scores for the above marker would be 152, 156, or so on. If those values are seen, they were changed to 151 and 157

3.4 Data Analysis

For general data analysis R (version 3.1.2; R core Team, 2015) software was used. The package '*Poppr*' version 1.1.2 (Kamvar *et al*., 2014) with additional libraries of: Ape v3.1-4 (Paradis *et al*., 2004), Adegenet v 1.4-2 (Jombart 2008) and Pegas v 0.6 (Paradis et al. 2004) was used for phylogenetic analysis. The function 'aboot' was used for constructing trees with Nei's distance (Nei 1972, 1978), neighbor-joining (Saitou and Nei 1987), a sample of 5,000 bootstrap replicates and a cut off of 75%. Identification of multi-locus genotype (MLG) was performed with MLGsim2.0 (Stenberg *et al*., 2003). It calculated the probability of MLGs sampled arising by chance (*PSex*) from the probability of independent occurrence (*PGen*). The P-value for testing the significance of *PSex* for each MLG was estimated using 1000 simulations. *PSex* values that fell below a *P*-value of 0.001, was concluded that identical genotypes originated from the same genet (individuals developed from a zygote and is, therefore, a product of sexual reproduction).

The clonal richness was estimated using the following indices calculated by 'RClone' package ver. 1.0.2 (Arnaud-Haond et al., 2017): the clonal diversity index (R); the Simpson's diversity index modified for finite sample sizes (D); the complement of the slope of the Pareto distribution (Pareto index) of clonal membership and the Pielou evenness index (J). SSR loci characteristics i.e alleles per loci, heterozygosity and fixation index were determined using GenAIEx 6.5. In addition to this, the AMOVA and inbreeding coefficient values were generated by the same software (Peakall and Smouse 2012).

CHAPTER FOUR

RESULTS

4.0 The level of genetic variation of *Pgt* population in Kenya 2015

Despite an initial screening of 20 *Pgt* labelled SSR markers, only 12 moderately polymorphic loci were found suitable for the analysis of genetic structure. Markers Pgest173, PgtCAA93 and PgtCAA39 were discarded due to their low signal quality while markers: Pgest341, Pgest021, Pgest098, Pgest325 and Pgest353 were discarded because the re-scoring of the electropherograms would not be validated the second time. In addition to this, locus, Pgest059 and PgtCAT4.2 were dropped during the analysis due to their monomorphic in nature. On average 6.09% of the data were scored as missing for each locus. Locus Pgest353 and PgtCAA39 had the highest missing data with 29.57% and 22.04%, respectively (Fig 2). A summary of primer sequences and characteristics of 10 SSR markers used in analysis (Table 4.0).

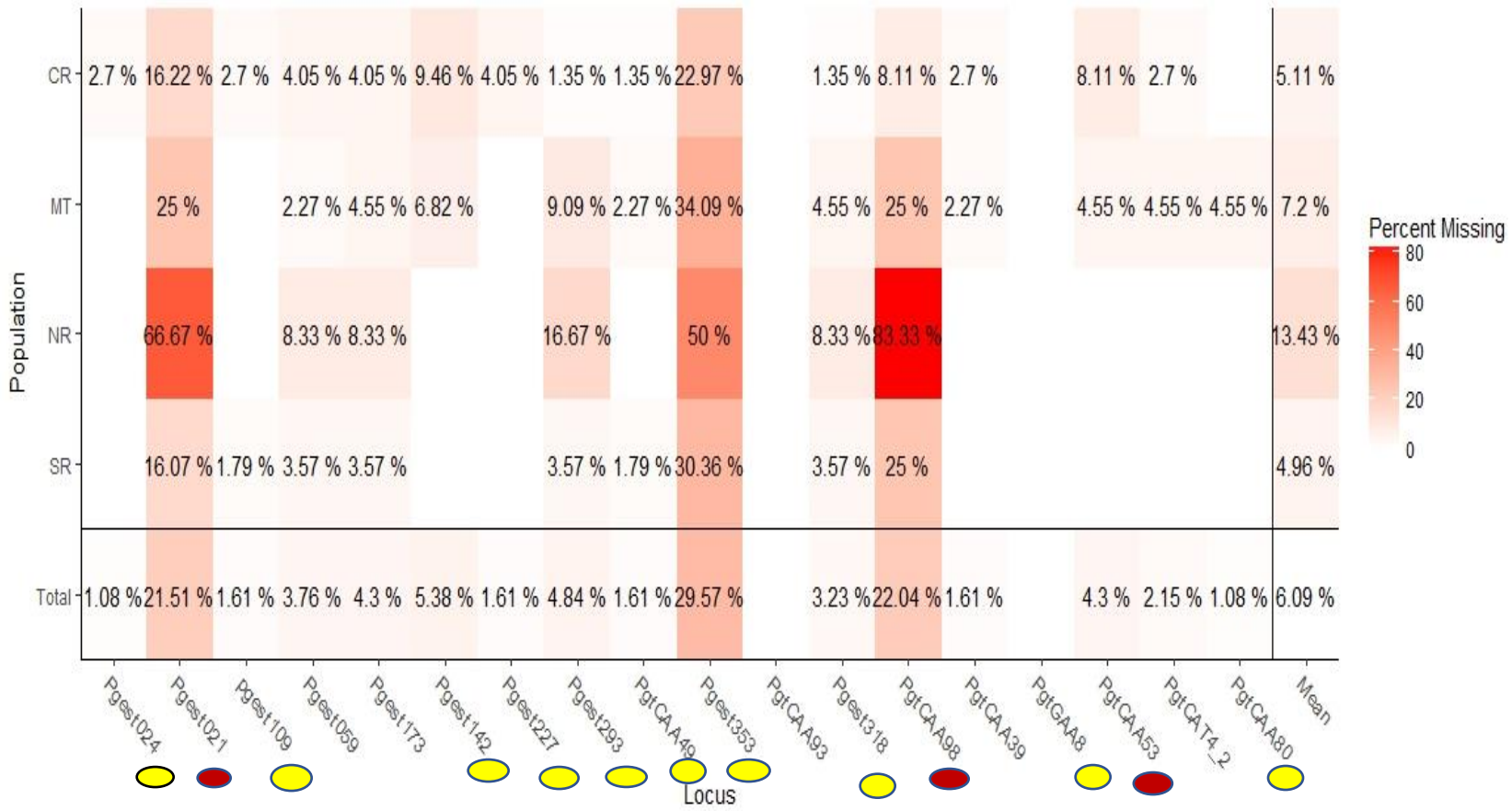


Fig 2. Percentage of missing data per locus and population of *Pgt* 2015 samples for populations NR (North Rift), SR (South Rift), MT (Mount Kenya) and CR (Central Rift). 10 markers retained and used for final analysis. Markers discarded without repeating the samples.

Table 4.0 Primer sequences and characteristics of the 10 SSR makers used in the genetic diversity study of Kenya *Pgt* samples 2015.

Locus	Repeat Motif	Primer	Primer Sequence	IRD	Supercontig	First base	Last base	amplicon size (bp)	Reference
Pgest142	AAC	142F-AAC	CCACCAACAAACCAACAAGA	700	17	917808	917827	213	Stoxen S, 2012
		142R	GATGGTGAAGTCCGGTATGG		17	917630	917649	198	
Pgest227	AAG	227F-AAG	CACACGTCTCGAGGAACAGA	700	33	496809	496828	202	Zhong et al. Phytopath 2009
		227R	CTCGTGGGATGAAGTCCATT		33	496991	497010	202	
Pgest293	ACC	293F-ACC	GAACCTTGGCCTGAGTGCTA	800	54	507344	507363	249	Zhong et al. Phytopath 2009
		293R	GCAGCCTACAGCAAGAATCC		54	507570	507589	246	
PgtGAA8.1	GAA	PgGAA8F1	GGATGATCGGTCAGTTGGTT	700	33	1385982	138617	212	Stoxen S, 2012
		GAA8R1	TGTCTGCCTGTCTGTCGAAC		33	138598	138617	212	
PgtCAA53	CAA	CAA53F1	AGGCTCAACACCACCCATAC	700	14	159139	159158	239	Jin et al. 2009
		CAA53R1	AGGAGGAGGTGAAGGGGATA		14	159358	159377	239	
Pgest024	AAG	21F-AAG	GTTTGCCTGATGATGGATGA	700	2	2065373	2065362	249	Stoxen S, 2012
		21R	CCGAATGCAGATTACCCTTG		2	2065144	2065163	249	
PgtCAA80	CAA	CAA80F1	GCCTCCAGACGAATGGTTTA	800	3	502155	502174	248	Stoxen S, 2012
		CAA80R1	TTGGTGATGATGATGGTTGG		3	502383	502402	248	
Pgest109	AGG	109F-AGG	CCATCCGATCATTTCTTCGT	700	13	54732	54751	176	Zhong et al. Phytopath 2009
		109R	CCGACCTTCTCTTGCTTCTG		13	54882	54901	170	
PgtCAA49	CAA	CAA49F1	TCGTCTGATCGTGAGAAACG	700	21	538804	538823	131	Stoxen S, 2012
		CAA49R1	GACGATTGCTGAGGATTGCT		21	538915	538934	131	
Pgest318	ATG	318F-ATG	ACAGACACTCCCGAGCTCAT	800	64	319750	319769	179	Stoxen S, 2012
		318R	GATGTCGGTCTTGGTCCACT		64	319514	319533	256	

SSR marker location information is from Stoxen S, 2012 based on the *Pgt* assembly (Duplessis et al, 2011).

In total we had 222 Kenya 2015 *Pgt* samples attempted to be used for the genetic studies scaled down from the original collection of 536 based on NanoDrop results. The conclusion made concerning the majority of samples, 314 in total was that they fell below the prescribed ratio of ~1.8 which is generally accepted as "pure DNA" even if the DNA quantity was low. After PCR amplification, re-binning of the data set was done so as to get the actual readings of the chromatographs. From this exercise, a total of 118 samples were further dropped as they were recorded as missing data thus reducing the samples to a total of 104 as it had no missing data thus used as the actual number in 2015 genetic studies. (Table 4.1).

Table 4.1 Number of samples used in the genotypic studies with their respective numbers and sampled fields.

Region	Field	No. Samples attempted to be genotyped	No. Samples genotyped
North Rift	Merewet	34	3
	Ray Farm (Sergoit)	40	15
South Rift	Narok Olopito	3	1
	Mbokishi	8	3
	Ngorengore	6	4
Central Rift	Kwa Ndusu Rongai	35	18
	Menengai	34	18
	Piave	8	1
Mt Kenya	Ngushishi-Maritati	27	23
	Marutu-Laikipia	27	18
	Total Samples	222	104

From the 10 *Pgt* SSR loci examined, allele frequency distribution ranged from 2 to 4 microsatellite alleles per locus and an average of 3.10 per locus. Locus Pgest318 had the highest Simpson diversity (0.74) while locus PgtCAA53 and Pgest318 had the most evenly distributed alleles (0.98) (Table 4.2)

The observed heterozygosity (H_o) for each of the 10 SSR loci were significantly different ($P < 0.001$) than the expected heterozygosity (H_e) with exception of loci Pgest024. The high values of observed heterozygosity (H_o) would mean there is mixing of the *Pgt* population thus making it to be single. Fixation Index (F) ranging from (-0.083 to -0.765) were observed. In practical sense the expected range should be from 0 to a theoretical maximum of 1. The former means no differentiation between the overall population and its subpopulation while the latter means highly differentiated overall population. From the data, the Fixation index (F) indicated that the *Pgt* population had low genetic variation. This would be further be validated by the number of effective alleles (N_e) ranging from 1.80-3.04 and Information index (I) in the range of 0.52-1.11 (Table 4.2)

Table 4.2. SSR marker and locus characteristics for Kenyan *Puccinia graminis* f. sp. *tritici* single pustule samples collected in 2015.

Locus	Na	1-D	Ne	E.5	I	Ho	He	F
Pgest142	3	0.62	2.37	0.91	0.90	1.000***	0.57	-0.765
Pgest227	3	0.58	2.26	0.83	0.81	0.569***	0.5	-0.083
Pgest293	3	0.62	2.37	0.91	0.90	1.000***	0.57	-0.765
PgtGAA8.1	3	0.62	2.37	0.91	0.90	1.000***	0.57	-0.765
PgtCAA53	4	0.74	3.04	0.98	1.11	1.000***	0.64	-0.599
Pgest024	3	0.49	1.80	0.69	0.73	0.413	0.42	-0.160
PgtCAA80	3	0.62	2.34	0.91	0.90	1.000***	0.57	-0.765
Pgest109	2	0.33	1.63	0.74	0.52	0.569***	0.34	-0.472
PgtCAA49	2	0.33	1.63	0.74	0.52	0.569***	0.36	-0.472
Pgest318	4	0.74	3.04	0.98	1.11	1.000***	0.64	-0.599

N_a = number of alleles observed; 1-D = Simpson index; N_e = number of effective alleles; E.5 = Evenness; I = Information index; H_O = observed heterozygosity; H_e = expected heterozygosity; F = Fixation Index
*** = $P < 0.001$

Neighbour joining (NJ) tree analysis was conducted so as to understand the genetic relationship between, a dataset containing 104 samples and two reference *Pgt* isolates representing Ug99 race group genotypes (clade I; 04KEN156/4 TTKSK and clade IV-B; 13ETH43_1 TKTTF). A total of five SSR-MLGs were identified. A total of 32 samples with inclusion of two reference isolates were included in drawing the neighbour joining tree. Samples were chosen based on distribution of SSR-MLGs from each wheat growing region. A total of three samples were chosen from each growing region. The analysed 32 genotyped samples collections were clustered with Clade I; 04KEN156/4 (race TTKSK) and Clade IV-B (race TKTTF) (Figure 3).

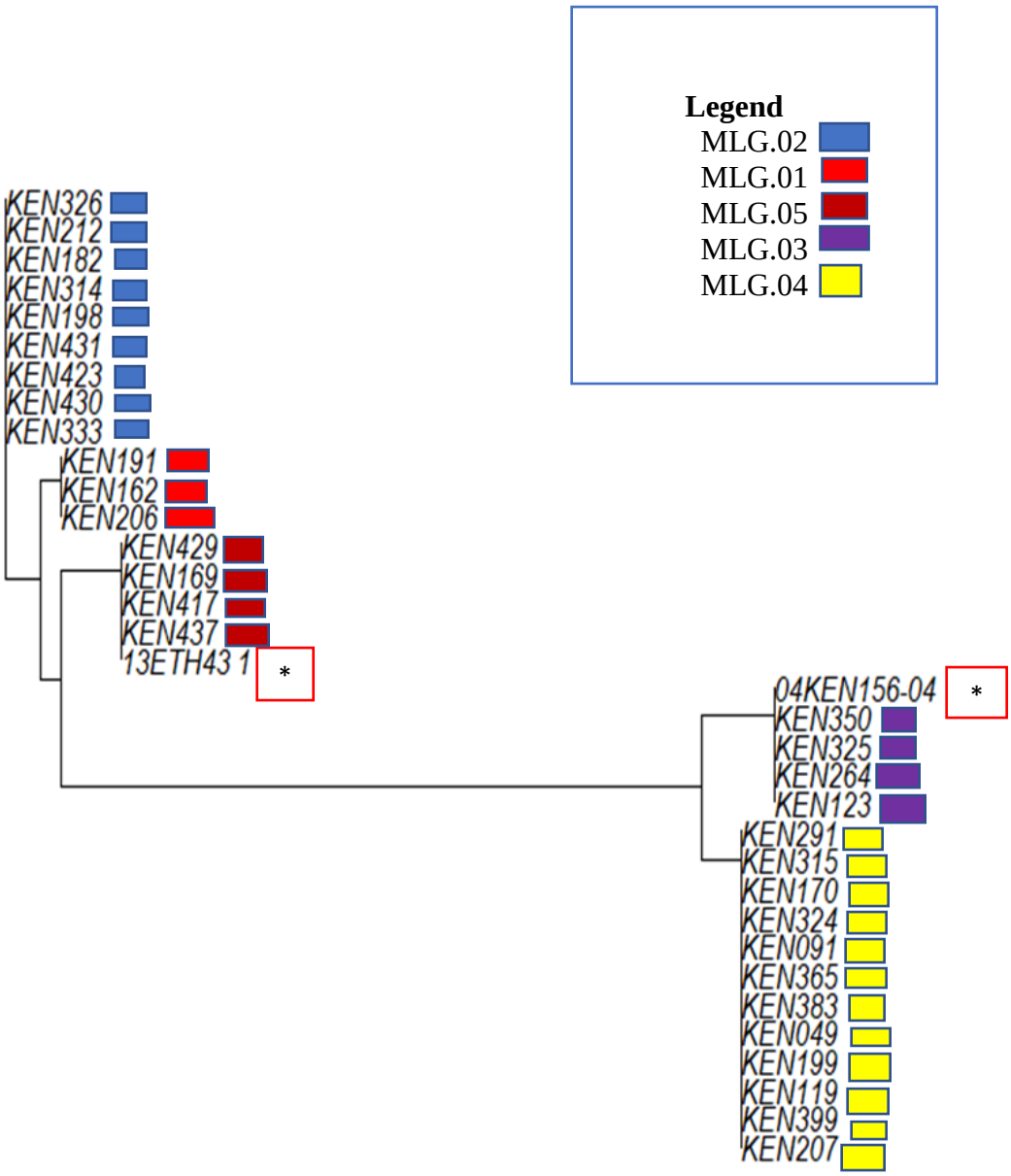


Fig 3. Neighbour-joining phylogenetic tree of 32 selected Kenyan *Puccinia graminis* f. sp. *tritici* samples based on 10 SSR markers. Two references isolates were included in analysis (*).

It was noted that samples having genotype SSR-MLGs.04 and SSR-MLGs.03 were clustered in clade I (race TTKSK and Ug99 race group) while SSR-MLG.01, SSR-MLG.02 and SSR-MLGs.05 were clustered in clade IV-B (race TKTTF).

F_{ST}-statistics was conducted to measure the amount of genetic variation among populations (wheat growing regions) to the overall total population (Kenya) with emphasis on the F'_{ST} and *D* estimates based on G''_{ST}. F'_{ST} and D were 0.119 and 0.184 respectively indicating no genetic differentiation as they were skewed towards zero. This would be further be validated by G''_{ST} which was 0.299 indicating low gene flow and mutation rate as the figure was close to zero. These figures were also relatively significantly different from ($p > 0.001$) *P* values based on the 9,999 permutations. The average inbreeding coefficient (F_{IS}) was a strong negative value of -0.56. This would support the suggestion that there is no genetic variation differences between wheat growing regions thus no geographical separation (Table 4.5).

Table 4.5. **Analysis of *F*st and related measures for *Puccinia graminis* f. sp. *tritici* field samples among populations and individuals in Kenya 2015 samples.**

Statistics	Value	*P*(rand >=data)
*F*is	-0.56	1.00
*F*st	0.119	0.001
*G*st	0.110	0.001
G"st	0.299	0.001
Dest (*D*)	0.184	0.001

P values based on 9,999 permutations. *F*is = Inbreeding coefficient within individuals *F*st = measure of population differentiation due to genetic structure, *G*st = Analog of Fst, adjusted for bias, *G*"st = Hedrick's standardized Gst, Dest = Jost's estimate of differentiation

Analysis of molecular variance (AMOVA) showed that the majority of the variation occurred within the samples (91%) rather than between regions (9%) further validating the

presence of no *Pgt* diversity and possibility of variation arising from asexual mutation. (Table 4.6)

Table 4.6. Analysis of molecular variance results for *Puccinia graminis* f. sp. *tritici* field samples among populations and individuals in Kenya 2015 samples.

Source	df	SS	MS	Est Var	%
Among Population	3	57.750	19.250	0.380	9%
Among samples	100	130.004	1.300	0.000	0%
Within samples	104	406.500	3.909	3.909	91%
Total	207	594.255		4.288	100%

df = degrees of freedom: SS = Sum of Squares: MS = Mean Squares: Est Var = Estimated Variance

Further analysis was done to examine the distribution of 5 SSR-MLGs in the field levels in relation to the major wheat crop variety grown in those farms. Mount Kenya, Magutu Laikipia (ML) farm had major and one dominant genotype SSR-MLG.04 while Ngushishi-Maritati (NM) had mixture of genotypes SSR-MLG.04,02,05 and 03. In Central Rift Kenya, farms Ndusu Rongai (KNR) and Piave (P) had dominant genotype SSR-MLG.02 while farm Menengai (ME) had mixtures of genotypes SSR-MLGs.04, 03 and 02. North Rift farms; Ray farm (RF) and Merewet (MW) farm had mixture of genotypes with SSR-MLGs.02 and 01 being the dominant ones. Two farms in the South Rift: Mbokishi (MB) and Narok Olopito (NO) had SSR-MLGs.02 being the major genotype with Ngorengore (NG) farm having mixtures of genotypes SSR-MLGs.02 and 01. (**Table 4.7)**

It would also be noted that from the total samples collected, frequency of genotype SSR-MLG.04 was in every farm with exception of Mount Kenya region with specificity to Magutu Laikipia farm while genotype SSR-MLG.01and SSR-MLG.03 frequency was low with only being recorded in two farms (**Table 4.7**)

Table 4.7 Distribution of five simple sequence repeats-multiple locus genotype (SSR-MLGs) between fields for 104 Kenyan *Puccinia graminis* f. sp. *tritici* samples of 2015 based on 10 *Pgt* single sequence repeats loci.

Region	Field	Wheat Cultivar	No of Samples	MLGs				
				MLG. 01	MLG.02	MLG.03	MLG.04	MLG.05
North Rift	Ray Farm	Robin	15	0	7 (47%)	0	6 (40%)	2 (13%)
North Rift	Merewet	Robin	3	0	0	0	2 (67%)	1 (33%)
South Rift	Ngorengore	K. Hawk 12	4	0	0	0	3 (75%)	1 (25%)
Mount Kenya	Ngushishi Maritati	Robin	23	3 (13%)	15(65%)	1 (4%)	4 (17%)	0
Central Rift	Menengai	Robin	18	0	12(67%)	4 (22%)	2 (11%)	0
Central Rift	Kwa Ndusu Rongai	Robin	18	0	0	0	18(100%)	0
Central Rift	Piave	Robin	1	0	0	0	1(100%)	0
South Rift	Mbokishi	Robin	3	0	0	0	3(100%)	0
South Rift	Narok Olopito	K. Hawk 12	1	0	0	0	1 (100%)	0
Mount Kenya	Magutu Laikipia	Robin	18	1(6%)	17(94%)	0	0	0

4.1 Genetic variation of *Pgt* population in Kenya 2015 brought about by asexual mutation and migration forces.

In order to know which MLGs are likely to be clones, MLGsim program was used. It uses simulation approach and calculates significance values for the likelihood of the observed number of MLGs in a specific sample from a population. It confirmed *psex* values from each SSR-MLGs to be significantly different (P <0.001) than the simulated *psex* (1.60E-08) meaning that they likely belong to the same genet. In total 5 SSR-MLGs were clonal. This would confirm that the SSR-MLGs generated are not as a result of sexual recombination rather than by asexual mutation (Table 4.8)

Table 4.8. **The *P-values* testing the significance of *PSex* for each SSR-MLG estimated using 1000 simulations.**

MLGs	n	*PSex*	Significance	Level	PGen
MLG.01	3	2.33E-10	0	***	1.09E-05
MLG.02	52	2.33E-15	0	***	8.18E-06
MLG.03	5	2.11E-15	0	***	7.41E-08
MLG.04	40	0	0	***	7.03E-08
MLG.05	4	1.11E-15	0	***	6.24E-10

Simulated *PSex* value=1.60E-08, significance *PValue* (0.001); MLG= Multi-locus genotype;
n= Number of occurrences

Clonal diversity (R) test was done to provide information about dynamics of the spread of the clones. It ranged from 0.075 to 0.143 with South Rift having the highest index while Central Rift recorded the least. Clonal diversity was lowest in Central Rift because of the high number of ramets per genet. This data was a reflection of a possible diversity based on the regions sample size given the fact that collections of samples was not evenly distributed. The diversity of the clones was measured using Simpson's diversity index which ranged from 0.25 to 0.660. The theoretical range is between 0 and 1 with the former representing infinite diversity and the latter no diversity. The evenness index which measures the closeness of the clones ranged from 0.493 to 0.934 which would support that the population has low diversity. The pareto index values ranged from 0.064 to 0.609 with Mount Kenya having the least value while North Rift having the highest value (Table 4.9). Strong negative values of F*is* a parameter of individuals that develop from a genet or another ramet as a result of asexual reproduction was recorded in the four regions, indicating that there was low inbreeding of individuals within the subpopulation. This variation would confirm the generation of SSR-MLGs being as a result of asexual reproduction on the wheat varieties grown in those regions.

Table 4.9. Gene and genotype diversity indices for *Pgt* collection of wheat growing regions in Kenya

Region	N	nb_B	R	D'	J	Pareto_Index	Fis	Fis_WR
Mount Kenya	41	4	0.075	0.345	0.493	0.064	-0.529	-0.631
Central Rift	37	3	0.056	0.577	0.844	0.308	-0.414	-0.120
North Rift	18	3	0.118	0.660	0.934	0.609	-0.389	-0.208
South Rift	8	2	0.143	0.25	0.544	0.069	-0.069	-0.364

N= number of samples; nb_B= number of MLG; R= genotypic richness/clonal diversity; D'= Probability of encountering distinct MLG when randomly taking 2 units in population; J'=clonal heterogeneity equitability; Pareto_index= distribution of clonal size in the population; Fis=on ramets if diploid data is product of asexual reproduction; Fis_WR,=on genets if diploid data is product of sexual reproduction.

4.2 Challenges associated with analysing the molecular SSR data

Challenges associated with my data analysis included the large amount of missing data from the collected regions. This would be greatly be attributed to samples having low DNA quality and quantity which lead to elimination of the first batch of sets of samples before running the SSRs. The second round of elimination of the samples was done on amplified PCR products which showed poor performance in the SSRs reactions. After the first two steps of elimination, some of the samples still recorded missing data even after the reactions were repeated twice. One of the possible explanations for this is that some of the samples had null alleles as the reaction failed to produce a visible product due to mutation at a priming site. In addition to this the collected samples were single-uredinial pustules compared to the reference isolates used which were multiple-uredinial pustules and no null alleles recorded in their reaction.

During scoring, part of the data had low signal quality especially when reading the peaks (electropherograms). This would make the patterns interpretation difficult especially in the case of reading adjacent -allele heterozygotes at a specific locus which have a dinucleotide motif repeat. In case of larger alleles, these heterozygotes can be scored as homozygotes leading to mistyping that leads to larger alleles biasedness and reduction of observed heterozygosity. Generally, these low signal electropherograms would be shown by stuttering patterns as some loci would produce stutter bands due to slipping by *Taq* polymerase (Figure 4).

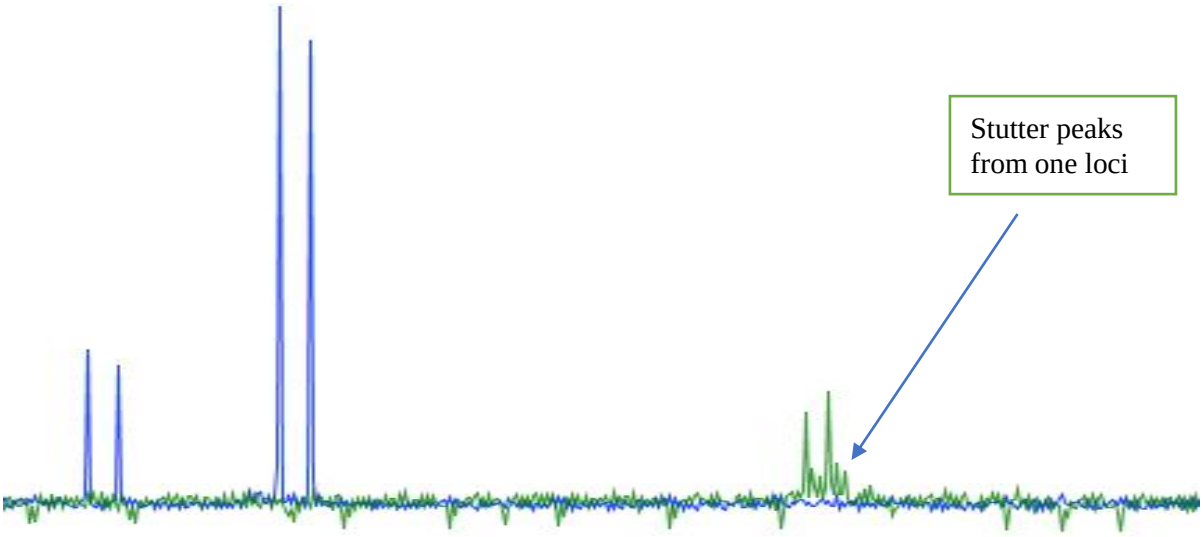


Fig 4. Schematic diagram showing two loci one of them having stutter bands

To go about this challenge, the reading of the electropherogram was taken from the two highest peaks of each loci.

Secondly, while scoring the data, some samples contained three similarly (electropherogram) sized peak. This at times lead to large allele drop as there was potential scoring error of allele and genotype frequencies biasedness. This challenge was sorted out by setting the required/known peak size for each and every marker before scoring. This made scoring easy as the amplified three peaks either small or large would de differentiated based on known set peak sizes (Fig 5).

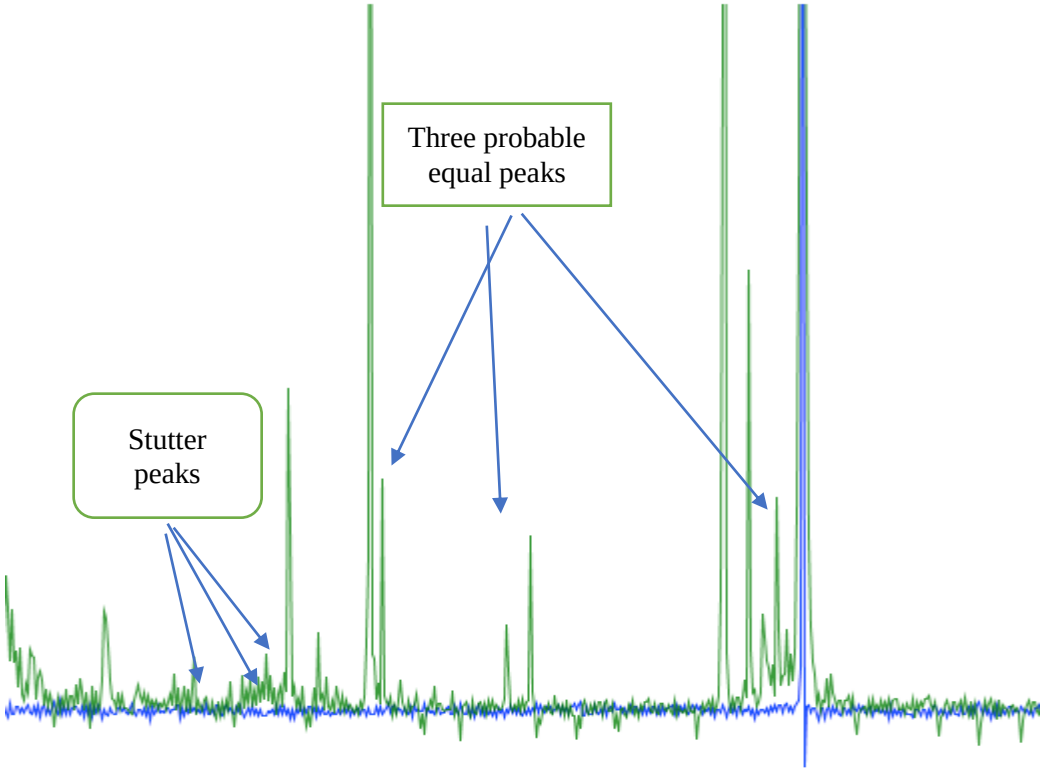


Fig 5. Schematic diagram showing one loci having three similar peaks (green).

CHAPTER FIVE

DISCUSSIONS

The future breeding programmes of any crop partially depends on understanding of their relevant pathogens' genetic variation and co-existence between the host and the pathogen. One of the crops, wheat has attracted interested since 2004 in terms of breeding resistant varieties against *Pgt* for future generation in Kenya and relevant countries having the known Ug99 race groups. The programme associated with breeding for resistant varieties have been improved through continuous monitoring of fungal pathogens with emphasis on *Pgt* genetic diversity (Singh *et al*., 2015). The dynamic nature of *Pgt* in Kenya has highlighted 10 variants within the Ug99 race groups out of the total 13 known variants. Currently, Kenya has been pinpointed as the most diverse country in terms of Ug99 race group represented by clade I ranging over a span of ten years (2004 to 2014). Recently there is entry of clade IV-B races in Kenya *Pgt* population expanding the number of stem rust clades from one to two which currently exists. (Newcomb *et al*., 2016; Pablo D. Olivera and Yue Jin, 2018).

This study was a follow up study of 2011 *Pgt* population in Kenya through the utilization of simple sequence repeats markers (SSR) to examine genetic structure of *Pgt* (Wanyera *et al*., 2017 in press). The results indicated that the *Pgt* population is single as there was no geographic population structure. In addition to this there was moderate level of diversity with the two dominant genotypes MLG.75 and MLG.47 representing Ug99 race group. This would be further be supported with Analysis of Molecular Variance (AMOVA) which showed majority of variation occurred within samples (95.5%) rather than between regions (1.2%).

Results from this current study of 2015 samples revealed the Kenya *Pgt* to be single with no evidence of geographical sub-structure. This would be validated by Analysis of molecular variance (AMOVA) which showed majority of variation occurring within samples (91%) rather than (9%) between regions. The strong negative values of fixation index (F) in the range of -0.083 to -0.765 indicated no overall differentiation between the samples collected from the four wheat growing regions.

The non-existence of clear geographical sub-structure would further be supported by F_{ST}-statistics data, which had a strong negative value of -0.56 from the average inbreeding coefficient (F_{IS}) suggesting no genetic variation even though we had two clades: clade I (Ug99 race group) and clade IV-B (race TTTTF/TKTTF). This slightly deviated from the previous research work done on the *Puccinia striformis* f.sp *tritici* (PST) pathogen which showed existence of geographical sub-division at regional level due to capability of the long-distance colonization of the pathogen (Hovmøller *et al*., 2011).

A total of five SSR-MLG genotype was found with one dominant genotype SSR-MLG.04 found in all four sampled wheat growing regions representing 50% of the population. This genotype (SSR-MLG.04) belonging to Clade I (Ug99 race group) was found in collections from 9 wheat fields sampled out of the 10 in total. From the neighbour joining tree data, there was indication of two clades with genotypes SSR-MLG.01,SSR-MLG.02 and SSR-MLG.05 represented Clade IV-B (race TKTTF/TTTTF) while genotypes SSR-MLG.04 and SSR-MLG.03 represented Clade I (Ug99 race group).This was a deviation from 2011 data which stated that the Kenya *Pgt* population to be consisting of genotypes representing clade I (Ug99 race group) only (Wanyera *et al*., 2017 in press).

The alternative hypothesis for the observed low genetic diversity of the 2015 Kenyan population is as a result of the migration and accumulation of an asexual mutations. The

presence of SSR-MLG.01 and SSR-MLG.02 both representing Clade IV-B which was originally found in Ethiopia would suggest that the spores would have been transported via wind. Clade IV-B (race TKTTF/TTTTF) was reported in Kenya 2015 through phenotyping with emphasis on race TTTTF. In 2017, race TKTTF and TTTTF was reported from samples collected (Pablo D. Olivera and Yue Jin, 2018). This data would suggest that the Clade IV-B has been in Kenya since 2015. After the epidemic of wheat in southern Ethiopia in the late 2013 growing season caused by race TKTTF, the collected d-samples (DNA samples) in 2014 recorded 48% of samples belonging to this race (Clade IV-A 1.8%; IV-B 40%). In 2015 95% of the d-samples collected and analysed from Ethiopia belonged to clade IV-B (Szabo 2018). This would suggest that the genotypes migrated from Ethiopia where its frequency was high to neighbouring country Kenya.

This would be supported by the previously done studies in Ethiopia that conceptualized modelling known as U.K Met Office Numerical Atmospheric-dispersion Modeling Environment (NAME) model to predict where the spores would go (Jones *et al*., 2007; Olivera *et al*., 2015). This was also consistent with previous done studies on PST that would suggest the capability of the pathogen to colonize a region through long distance (Brown and Hovmøller 2002).

From the present results, it would be noted that the host variety contributed in the selective force of the *Pgt* population in relation to favouring the pathogen that has virulence to the host *Sr* genes. In 2015 two wheat host varieties Robin (*Sr*Tmp) and K.Hawk 12 (*Sr* 2, *Sr*Tmp) were widely grown. Out of the ten farms sampled, South Rift-Narok Olopito and South Rift-Ngorengore had varieties K.Hawk 12. The other remaining eight farms had varieties Robin. Both varieties have resistant gene *Sr*Tmp which is susceptible to *Pgt* race TKTTF (Clade IV-B) (Olivera *et al*., 2015). These findings would be supported by

previous work done on PST reported between North and South of France (Enjalbert *at al*., 2005) which was explained that the stability of the PST population subdivision over years to be attributed by difference in host resistance gene deployment and adaptation of the pathogen to the localized temperature (Mboup *et al*., 2012).

The asexual mutation of the population was observed with the significantly different *psex* values from each 5 SSR-MLGs ($P < 0.001$) than the simulated *psex* (1.60E-08). In addition to this, the 5 SSR-MLGs originated from the same genet. This would be validated by strong negative values of F*is* ranging from -0.069 to -0.529 recorded from the four regions signifying low inbreeding. Further validation was shown from the 10 SSR loci showing observed heterozygosity (H_o) being significantly ($P < 0.001$) different from expected heterozygosity (H_e) with exception of loci Pgest024.

Genotypic diversity was low at all locations where sampling was done with the minimum value of 0.25 recorded in South Rift. Generally, the Simpson's diversity index in the range of 0.25 to 0.660 and evenness index values (0.493 to 0.934) were below one. These results are not consistent with the existence of recombination in wheat yellow rust pathogen population studies where microsatellites analysed showed high levels of genotypic diversity no significant difference between observed and expected heterozygosity (Ali, 2012). The low presence of *Berberis* spp the alternate host for sexual cycle of *Pgt* found around Mount Kenya region would also contribute to low genotypic diversity. This would be further supported with studies done in Asia which showed existence of high genotypic diversity and sexual cycle of wheat yellow rust pathogen in the presence of the abundance of *Berberis* spp (Ali, 2012). From the current findings, 2015 Kenya *Pgt* population has no sexual recombination signature despite the migration of genotype clade IV-B (race TKTTF/TTTTF) from Ethiopia to Kenya most probably via wind.

Given the fact that the regions shared the same multi-locus genotypes SSR-MLG. 02 and SSR-MLG.04, with exception of South Rift would suggest that the individuals are clones. The temporal maintenance of the *Pgt* especially in the non-*Berberies* zone, as a result of off-season over-summering, seems to be a source of the migrating genotypes.

Due to the diversity and wide distribution of the pathogen, national and international collaborations have been made so as to boost the surveillance of the wheat stem rust pathogen. Some of the collaborating institutions include: Global Rust Reference Center (GRRC) in Denmark and USDA-ARS Cereal Disease laboratory (USA). Fifteen African countries, Kenya and Ethiopia inclusive usually send *Pgt* infected wheat tissue for race phenotyping based on the method of Jin *et al*,.2008 to the GRRC. For the purposes of genotyping using the *Pgt* SSR and SNP markers the samples are usually shipped to USDA-ARS Cereal Disease laboratory (USA). The ultimate agenda of this yearly research work is to update the status of *Pgt* in the hot spot countries in terms of race distribution and diversity.

Prior to 2011 studies, little was known about Kenya's regional distribution of *Pgt* races and members of this genetic lineage. With the combined utilization of phenotypic and genotypic data, understanding of the distributions and frequencies of *Pgt* races in Kenya as well as its neighbouring countries has boosted the pathogen surveillance (Olivera *et al*., 2015; Newcomb *et al*., 2017; Wanyera *et al*., 2017 in press). The data is expected to boost the durable breeding resistance efforts against the pathogen with an aim to increase food security especially through wheat crop (Chakraborty and Newton 2011). In addition to this, the Kenya breeding program is expected to focus on developing and releasing cultivars with *Sr* genes that show resistance to broad spectrum of virulence with emphasis on clade I (Ug99 race group) and clade IV-B (race TKTTF).

CHAPTER SIX

GENERAL CONCLUSION AND RECOMMENDATIONS

6.0 Conclusions

This study confirmed the existence of single *Pgt* population in Kenya made up of two phylogenetic clades, clade I (Ug99 race group) and clade IV-B (race TKTTF/TTTTF). It was an indication that the *Pgt* population in Kenya shifted in composition from only being based on Clade I (Ug99 race group) in 2011 with the introduction of Clade IV-B (race TKTTF/TTTTF) in 2015. In addition to this the genetic variation in *Pgt* population in Kenya was low among samples collected from the four main regions: North Rift, South Rift, Central Rift and Mount Kenya and arises from asexual mutation and migration from neighboring countries and regions.

6.1 Recommendations

1. Implementing diagnostic tools for rapid detection of stem rust races through utilization of DNA fingerprinting tools (SSR markers) has proved to be robust, timely and precise.
2. Understanding the pathogenic diversity of the stem rust races will enable deployment of resistance sources in wheat breeding effective against races with emphasis on clade I (Ug99 race group) and clade IV-B (race TKTTF/TTTTF).

REFERENCES

Abdel-Mawgoud, AM., Le´pine F and De´ziel, E. (2010) Rhamnolipids: diversity of structures, microbial origins, and roles. *Appl Microbiol Biotechnol* 86:1323–1336.

Agrios, G.N. (2005). *Plant Pathology*. 5th edition: Diseases Caused by Basidiomycetes (pp273-383). Elsevier Academic Press, New York.

Aime MC, Matheny PB, Henk DA, Frieders EM, Nilsson RH, Piepenbring M, McLaughlin DJ, Szabo LJ, Begerow D, Sampaio, J.P., Bauer, R., Weiss, M., Oberwinkler, F. and Hibbett, D. (2006). An overview of the higher-level classification of *Pucciniomycotina* based on combined analyses of nuclear large and small subunit rDNA sequences. *Mycologia* 98: 895–905.

Ali S (2012) *Population biology and invasion history of Puccinia striformis f.sp. tritici at worldwide and local scale*, Université Paris-Sud 11.

Akhunov, E., Nicolet, C. and Dvorak, J. (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina Golden Gate assay. *Theor. Appl. Genet*.119:507–517.

Arnaud-Haond, S., Bailleul, D., and Stoeckel, S (2016). RClone: a package to identify Multilocus Clonal Lineages and handle of clonal data sets in R. *Methods in Ecology and Evolution* 7, 966-970.

Atallah, Z.K., Maruthachalam, K., Toit, L.d., Koike, S.T., Davis, R.M., Klosterman, S.J., Hayes, R.J., and Subbarao, K.V (2010). Population analyses of the vascular plant pathogen *Verticillium dahliae* detect recombination and transcontinental gene flow. *Fungal Genet.* Biol. 47:416–22

Bateson, W., 1909 Mendel's Principles of Heredity. *Cambridge University Press*, Cambridge, UK.

Borlaug, N.E. (1972) A cereal breeder and ex-forester's evaluation of the progress and problems involved in breeding rust resistant forest trees. *In Proc. of a NATO-IUFRO Advanced Study Institute, pp.* 615-642. Aug. 17-24, 1969 "Moderator's Summary". Biology of Rust Resistance in Forest Trees. USDA Forest Service Misc. Publication 1221.

Boshoff W H P, Pretorius Z A, van Niekerk B D and Komen J S. (2002) First report of virulence in *Puccinia graminis* f. sp. *tritici* to wheat stem rust resistance genes *Sr8b* and *Sr38* in South Africa. *Plant Disease* 86:922.

Boehm, E.W.A., Wenstrom, J.C., McLaughlin, D.J., Szabo, L.J., Roelfs, A.P. and Bushnell, W.R. (1992) An ultrastructural pachytene karyotype for *Puccinia graminis* f.sp. *tritici*. Can. J. Bot. 70, 401-413.

Brown JKM, and Hovmøller MS. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* 2002;297:537–41.

Chakraborty, S., and Newton, C. 2011. Climate change, plant diseases and food security: An overview. *Plant Pathology*. 60:2-14

Chester K S. (1946) The nature and prevention of the cereal rusts as exemplified in the leaf rust of wheat. Waltham, MA: *Chronica* Botanica.

Cuomo CA and Birren BW (2010) The fungal genome initiative and lessons learned from genome sequencing. Methods *Enzymol* **470**:833–855.

Cuomo, C. A., Bakkeren, G., Khalil, H. B., Panwar, V., Joly, D., Linning, R., et al. (2017). Comparative analysis highlights variable genome content of wheat rusts and divergence of the mating loci. G3: *Genes-Genomes-Genetics*, 7(2), 361–376.

David Hodson (2011) www.rusttracker.org

Doyle, J. J. and J. L. Doyle. (1990). A rapid total DNA preparation procedure for fresh plant tissue. *Focus* 12:13-15.

Duplessis S, Cuomo C A, Lin Y, Aerts A, Tisserant E, Veneault-Fourrey C, Joly D L, Hacquard S, Amselem J, Cantarel B L, Chiu R, Coutinho P M, Feau N, Field M, Frey P, Gelhaye E, Goldberg J, Grabherr M G, Kodira C D, Kohler A, Kues U, Lindquist E A, Lucas S M, Mago R, Mauceli E, Morin E, Murat C. Pangilinan J L, Park R, Pearson M, Quesneville H, Roucheir N, Sakthikumar S, Salamov A A, Schmutz J, Selles B, Shapiro H, Tanguay P, Tuskan G A, Henrissat B, Van de Peer Y, Rouze P, Ellis J G, Dodds P N, Schein J E, Zhong S, Hamelin R C, Grigoriev I V, Szabo L J and Martin F. (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Science* 108:9166-9171.

Ellis, J.G., Lagudah, E.S., Spielmeyer, W., and Dodds, P.N. (2014). The past, present, and the future of breeding rust resistant wheat. Front. Plant Sci. 5:1-13.

Enjalbert J, Duan. X, Leconte M, Hovmøller, M.S and de Vallavieille, Pope C (2005). Genetic evidence of local adaptation of wheat yellow rust (*Puccinia striiform*is f.sp *tritici*) *Molecular Ecology*, 14, 2065-2073.

Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE. (2003). Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* 93:738–46

Grünwald N.J, Kamvar,Z.N and Everhart, S.E (2016). Grunwaldlab/Population Genetics in R: First release [Data set]. Zenodo. http://doi.org/10.5281/zenodo.160588.

Gupta P. K., and R.K. Varshei (2000). The development and use of Microsatellite Markers for Genetic Analysis and Plant Breeding with Emphasis on Bread Wheat, *Euphytica*, 113, 163-185.

Hovmøller, M.S, Sørensen, C.K, Walter. S and Justesen, A.F. (2011). Diversity of *Puccinia striiformis* on cereals and grasses. *Annual review of Phytopathology*, 49, 197-217.

James B. Reid and John J. Ross (2011). Mendel's Genes: Toward a Full Molecular Characterization. GENETICS *September 1, 2011 vol. 189 no. 1 3-10; https://doi.org/10.1534/genetics.111.132118*

Jin Y, Szabo L J, Pretorius Z A, Singh R P, Ward R and Fetch T Jr. (2008) Detection of virulence to resistance gene *Sr24* within race TTKS of *Puccinia graminis* f. sp. *tritici*. *Plant Disease* 92:923-926.

Jin, Y., Szabo, L.J., Rouse, M.N., Fetch, T., Jr., Pretorius, Z.A., Wanyera, R., and Njau. P. (2009). Detection of virulence to resistance gene Sr36 within the TTKS race lineage of Puccinia graminis f. sp. tritici. *Plant Disease*. 93: 367-370

Johnson T. (1949) Intervarietal crosses in *Puccinia graminis*. *Canadian Journal of Research*, Section C. 27:45-65.

Jones A., Thomson D., Hort M., and Devenish B. (2007) The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. In: Borrego C., Norman AL. (eds) Air Pollution Modeling and Its Application XVII. *Springer,* Boston, MA

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.

Kamavar, Z.N., Tabima, J.F. and Grünwald, N.J. (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and /or sexual reproduction. *PeerJ*, 2,e281

Kamwaga J., Macharia G., Boyd L., Chiurugwi T., Midgley I., Canales C., Marcheselli M., Maina I., 2016. *Kenya Wheat Production Handbook*. Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya. ISBN 978-9966-30-025-6

Les J. Szabo Surveillance *Pgt*-core SNP assay. *BGRI Technical Workshop* Surveillance Meeting April 14th, 2018 Marrakech Morocco.

Keiper FJ, Hayden MJ, Park RF, Wellings CR (2003) Molecular genetic variability of Australian isolates of five cereal rust pathogens. *Mycological Research*, **107**, 545–556.

Keiper, F. J., Haque, M. S., Hayden, M. J., and Park, R. F. 2006. Genetic diversity in Australian populations of *Puccinia graminis* f. sp. *avenae*. *Phytopathology* 96:96-104.

Kumar A, Kumar S, Kumar R, Kumar V, Prasad L, Kumar N, Singh D. (2010). Identification of blast resistance expression in rice genotypes using Molecular markers (RAPD & SCAR). *Afri.J. of Biotech*. 09(24): 3501-3509.

Kubelik, A.R. & Szabo, L.J. Curr Genet (1995) 28: 384. https://doi.org/10.1007/BF00326438

Leonard K and Szabo L J. (2005) Stem rust of small grains and grasses caused by *Puccinia graminis*. Plant Molecular Pathology 6:99-111.

Martens, J.W (1978) Virulence dynamics in the wheat stem rust population in Kenya. *Plant Dis Reptr* 59;763-767.

Martin F, Aerts A, Ahrén D, Brun A, Danchin EG, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buée M, Brokstein P, Canbäck B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbé J, Lin YC, Legué V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculita-Hirzel H, Oudot-Le Secq MP, Peter M, Quesneville H, Rajashekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kües U, Lucas S, Van de Peer Y, Podila GK, Polle A, Pukkila PJ, Richardson PM, Rouzé P, Sanders IR, Stajich JE, Tunlid A, Tuskan G and Grigoriev IV (2008). The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis *Nature* **452**:88–92

Mboup M, Bahri B and Leconte M (2012). Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptations. *Evolutionary, Applications*, 5, 341-352. Doi:1111/j.1752-4571.2011.00228. x.

McIntosh R A, Wellings C R and Park R F. (1995) Wheat rusts: An atlas of resistance genes. London: *Kluwer Academic Publishers*.

Meirmans, P.G. and Van Tienderen, P.H (2004) GENOTYPE AND GENODIVE: two programs for the analysis of genetic diversity of asexual organism. *Molecular Ecology Notes*, 4, 792-794.

Mondal. K. Kalyan and Shanmugam (2013). Advancement in diagnosis of bacterial plant pathogens: An Overview. *Biotechnology and Molecular Biology* Review Vol 8 pp 1-11.

Newcomb, M., Olivera, P.D, Rouse, M.N., Szabo, L.J., Johnson, J., Gale, S., Luster, D.G., Wanyera, R., Macharia, G., Bhavani, S., Hodson, D., Patpour, M., Hovmøller, M.S, Fetch, T.G., and Jin, Y. (2016). Kenyan Isolates of *Puccinia graminis* f. sp. *tritici* from 2008 to 2014: Virulence to *SrTmp* in the Ug99 Race Group and Implications for Breeding Programs *Phytopathology*, 106(7), 729-736.

Njau PN, Jin Y, Huerta-Espino J, Keller B and Singh RP (2010): Identification and evaluation of sources of resistance to stem rust race Ug99 in wheat. In: *Plant Disease* 94:413-419

Nei, M. (1972). Genetic distance between populations. *Am. Nat*. 106: 283-292.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Olivera Firpo P.D., Newcomb M., Flath K., Sommerfeldt-Impe N., Szabo L.J., Carter M., Luster D.G.,and Jin Y. (2017) Characterization of *Puccinia graminis* f. sp. *tritici* isolates derived from an unusual wheat stem rust outbreak in Germany in 2013 *Plant Pathology*, 66 (8) , pp. 1258-1266.

Pablo D. Olivera and Yue Jin, Global Stem Rust Update. BGRI Technical Workshop Surveillance Meeting April 14th, 2018 Marrakech Morocco.

Park R, Fetch T, Hodson D, Jin Y, Nazari K, Prashar M, and Pretorius Z (2011). International surveillance of wheat rust pathogens: progress and challenges. *Euphytica* 179:109-117.

Park, PF (2007). Stem rust of wheat in Austarlia. *Aust J Agric Res* 58:558-566.

Park. R, Fetch. T, Hodson. D, Jin. Yue, Nazari. K, Prashar. M, and Pretorius. Z. (2011). International surveillance of wheat rust pathogens: progress and challenges. *Euphytica* **179**, 109–117.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE. Analysis of phylogenetics and evolution in R language. *Bioinformatics* 20: 419-420.

Patpour M., Hovmoller M., Justesen A. F., Newcomb M., Olivera Firpo P. D., Jin Y., Szabo, L.J., Hodson, D., Shahin, A.A., Wanyera, R., Habarurema, I., Wobibi, S. (2016). Emergence of virulence to *SrTmp* in the Ug99 race group of wheat stem rust, *Puccinia graminis* f. sp. *tritic*i, in Africa. *Plant Dis*. 100:522

Patpour, M., Hovmøller, M. S., Justesen, A. F., Newcomb, M., Olivera, P., Jin, Y., Szabo, L. J., Hodson, D., Shahin, A. A., Wanyera, R., Habarurema, I., and Wobibi, S. (2015). Emergence of virulence to *SrTmp* in the Ug99 race group of wheat stem rust, *Puccinia graminis* f. sp. *tritici*, in Africa. *Plant Dis*. 100(2),522

Pretorius Z A, Singh R P, Wagoire, W and Payne, T S. (2000) Detection of virulence to wheat stem rust resistance gene Sr31 in *Puccinia graminis* f. sp. *tritici* in Uganda. *Plant Diseases* 84:203

Pretorius, Z.A., Szabo, L.J., Boshoff, W.H.P., Herselman, L., and Visser, B. (2012). First report of a new TTKSF race of wheat stem rust (*Puccinia graminis* f.sp. *tritici*) in South Africa and Zimbabwe. *Plant Dis*. 96:590.

R Core Team (2015) *R*: *A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. *Online publication*. http://www.R-project.org/

Roelfs A P and Martens J W. (1988) An international system of nomenclature for *Puccinia graminis* f.sp. *tritici*. *Phytopathology* 78:526-533.

Roelfs A P. (1985) Wheat and rye stem rust. In the Cereal Rusts, Vol. 2 (Roelfs, A.P. and Bushnell, W.R., eds). Orlando, FL: *Academic Press,* pp 3-37.

Roelfs, A. P., Singh, R. P., and Saari, E. E. (1992). *Rust Diseases of Wheat*: Concepts and Methods of Disease Management. CIMMYT, Mexico, D.F.

Rouse, M.N., Talbert, L.E., Singh, D., and Sherman, J.D. 2014b. Complementary epistasis involving *Sr12* explains adult resistance to stem rust in Thatcher wheat (*Triticum aestivium* L.). *Theor. Appl.Genet*. 127:1549-1559.

Saitou, N., and Nei, M. 1987. The neighbor-joining method. A new method for reconstructing phylogenetic trees. *Mol. Bio. Evol*. 4:406-425.

Semenov M.A and Halford N.G. (2009). Identifying target traits and molecular mechanism for wheat breeding under a changing climate. *J. Exp. Bot*; 60:2791-2804

Singh, R.P., Hodson, D.P., Huerta-Espino, J., Jin, Y., Bhavani., Njau, P., Herrera-Foessel, J., Singh, P.K., Singh, S., and Govindan, V. (2011a). The Emergence of Ug99 Races of the Stem Rust Fungus is a threat to World Wheat Production Annual Review of *Phytopathology* Vol. 49: 465-481

Singh, R.P., Hodson, P., Jin, Y., Lagudah, E.S., Ayliffe, Bhavani, B., Rouse, M.M., Pretorius, Z.A., Szabo, L.J., Huerta-Espino, J., Basnet, B.R., Lan, C and Hovmøller, M.S. (2015). Emergence and Spread of New Races of Wheat Stem Rust Fungus: Continued Threat to Food Security and Prospects of Genetic Control**.** *Phytopathology* 105(7):872-84.

Singh, R.P., Huerta-Espino, J., Bhavani,S., Herrera-Foessel, S.A., Singh, D., Singh, P.K., Velu, G., Mason, R.E., Jin, Y., Njau, P., and Crossa, J. (2011b). Race non-specific resistance to rust disease in CIMMYT springs wheats. *Euphytica* 179:175-186.

Singh RP, William HM, Huerta-Espino J, Rosewarne G. Wheat rust in Asia: meeting the challenges with old and new technologies. In: New Directions for a Diverse Planet: Proceedings of the 4th International Crop Science Congress, Brisbane, Australia, 26 September–1 October 2004. 2004. ISBN 1 920842 20 9.

Skrede I, Carlsen T, Stensrud O, Kauserud H. (2012). Genome wide AFLP markers support cryptic species in Conphora (Boletales). *Fungal Biology*. 116:778-784.

Stakman E C and Levine M N. (1922) The determination of biologic forms of *Puccinia graminis* on *Triticum* spp. *Minnesota Agriculture Experimental Station Technical Bulletin 8 pp*.

Stenberg, P., Lundmark, M and Saura, A. (2003) mlgsim: a program for detecting clones using a simulation approach. *Molecular Ecology Notes*, 3, 329-331.

Stoxen S, 2012. *Population structure of* Puccinia graminis *f. sp.* tritici *in the United States*. St. Paul, USA, University of Minnesota, *M.S. thesis*. http://hdl.handle.net/11299/131092

Szabo, L.J (2007). Development of simple sequence repeats markers for the plant pathogenic rust fungus fungus, *Puccinia graminis*. *Mol. Ecol. Notes*, 7,92-94.

Szabo, Les.J. Surveillance *Pgt*-Core SNP assay. Global Stem Rust Update. *BGRI Technical Workshop* Surveillance Meeting April 14th, 2018 Marrakech Morocco.

Toome-Heller, M. (2016). Latest developments in the research of rust fungi and their allies (Pucciniomycotina). In D.-W. Li (Ed.), *Biology of microfungi* (pp. 147–168). Cham: *Springer International Publishing.*

Urban Z. (1967) The taxonomy of some European graminicolous rusts. *Ceska Mykologia* 21:12-16.

Visser, B., Herselman, L., Park, R. F., Karaoglum, H., Bender, C. M., and Pretorius, Z. A. (2011). Characterisation of two new *Puccinia graminis* f. sp. *tritici* races within the Ug99 lineage in South Africa. *Euphytica* 179, 119–127.

Visser B, Herselman L and Pertorius Z A. (2009) Genetic comparison of Ug99 with selected South African races of *Puccinia graminis* f. sp. *tritici*. *Molecular Plant Pathology* 10:213-222

Vurro M, Bonciani B and Vannacci G. (2010) Emerging infectious diseases of crop plants in developing countries: impacts on agriculture and socio-economic consequences. *Food Security* 2:113-132.

Wanyera, R., Wamalwa, M., Odemba, M., Wanga, H., Kinyanjui, P., Onyango, V., & Owuoche, J. (2016). *Management of wheat rusts at different growth stages* using Nativo 300 SC (trifloxystrobin 100g/L+ tebuconazole 200g/L) fungicide.

Wanyera R, Kinyua M G, Jin Y and Singh R P. (2006) The spread of stem rust caused by *Puccinia gramins* f. sp. *tritici*, with virulence on *Sr*31 in wheat in Eastern Africa. *Plant Disease* 90:113.

Wanyera, R., Kyalo, M., Wanjala, M., Harvey, J., and Szabo, L.J (2017). Genetic characterization of the 2011 wheat stem rust pathogen population in Kenya using SSR markers. *Plant pathology in press*.

Woo S L, Zoina A, Sorbo G D, Lorito M, Nanni B, Scala F, Noviello C. (1996). Characterization of *Fusarium oxysporum* f. sp. *phaseoli* by pathogenic races, VCGs, RFLPs and RAPD. *Phytopathol*. 86: 966-973.

Yang, H., Li, C., Lam, H.M., Clemements, J., Yan, G., and Zhao, S. (2015). Sequencing consolidates molecular markers with plant breeding practice. *Theor Appl Genet* 128;779-795.

Zambino, P. J and Szabo, L.J. (1993) Phylogenetic relationships of selected cereal and grass rusts based on *rDNA* sequence analysis. *Mycologia* 85:401–414.

Zambino, P. J., Kubelik, A. R., and Szabo, L. J. (2000). Gene action and linkage of avirulence genes to DNA markers in the rust fungus *Puccinia graminis*. *Phytopathology* 90:819-826.

Zhong S, Leng Y, Friesen T L, Faris J D, Szabo L J. (2009). Development and characterization of expressed sequence tag (EST)-derived microsatellite markers for the wheat stem rust fungus, *Puccinia graminis* f.sp. *tritici*. *Phytopathology* 99:282-289.

APPENDICES

Appendix I: Information on DNA quantity of 222 samples based on nanogram/microliter and 260/280 ratio

	SAMPLE ID	SAMPLED FARM	ng/ul	A260	A280	260/280	260/230
SOUTH RIFT	15KEN006	NAROK OLOPITO	44.68	0.894	0.426	2.1	2.27
	15KEN011	NAROK OLOPITO	65.45	1.309	0.93	1.41	0.38
	15KEN049	NAROK OLOPITO	59.25	1.185	0.675	1.76	0.63
	15KEN055	MBOKISHI	67.95	1.359	0.734	1.85	1.17
	15KEN060	MBOKISHI	73	1.46	0.826	1.77	0.95
	15KEN067	MBOKISHI	78.62	1.572	0.876	1.8	1.12
	15KEN073	MBOKISHI	63.85	1.277	0.753	1.7	0.82
	15KEN080	MBOKISHI	56.9	1.138	0.711	1.6	0.76
	15KEN084	MBOKISHI	162.97	3.259	2.729	1.19	0.51
	15KEN091	MBOKISHI	44	0.88	0.509	1.73	0.73
	15KEN119	NGORENGORE	18.88	0.378	0.234	1.62	1.97
	15KEN121	NGORENGORE	21.37	0.427	0.269	1.59	1.14
	15KEN123	NGORENGORE	23.01	0.46	0.253	1.82	1.53
	15KEN129	NGORENGORE	38.73	0.775	0.451	1.72	1.05
	15KEN130	NGORENGORE	25.68	0.514	0.291	1.76	1.21
	15KEN131	NGORENGORE	54.96	1.099	0.613	1.79	1.67
MT KENYA	15KEN156	NGUSHISHI-MARITATI	52.14	1.043	0.53	1.97	0.96
	15KEN161	NGUSHISHI-MARITATI	56.66	1.133	0.591	1.92	0.95
	15KEN162	NGUSHISHI-MARITATI	82.1	1.642	0.865	1.9	0.89
	15KEN164	NGUSHISHI-MARITATI	48.16	0.963	0.539	1.79	0.87
	15KEN165	NGUSHISHI-MARITATI	53.56	1.071	0.575	1.86	0.83
	15KEN169	NGUSHISHI-MARITATI	50.03	1.001	0.584	1.71	0.87
	15KEN170	NGUSHISHI-MARITATI	73.18	1.464	0.833	1.76	0.93
	15KEN171	NGUSHISHI-MARITATI	38.95	0.779	0.436	1.79	0.98
	15KEN172	NGUSHISHI-MARITATI	51.88	1.038	0.578	1.79	0.9
	15KEN173	NGUSHISHI-MARITATI	83.32	1.666	0.958	1.74	1.02
	15KEN174	NGUSHISHI-MARITATI	80.74	1.615	0.909	1.78	1.11
	15KEN175	NGUSHISHI-MARITATI	34.37	0.687	0.383	1.8	0.96
	15KEN176	NGUSHISHI-MARITATI	29.41	0.588	0.364	1.61	0.69
	15KEN178	NGUSHISHI-MARITATI	63.92	1.278	0.739	1.73	0.94

	15KEN179	NGUSHISHI-MARITATI	44.64	0.893	0.462	1.93	1.17
	15KEN180	NGUSHISHI-MARITATI	61.29	1.226	0.712	1.72	0.78
	15KEN182	NGUSHISHI-MARITATI	35.64	0.713	0.434	1.64	0.85
	15KEN190	NGUSHISHI-MARITATI	96.34	1.927	1.067	1.81	1.11
	15KEN191	NGUSHISHI-MARITATI	371.79	7.436	4.209	1.77	1.55
Appendix I	15KEN193	NGUSHISHI-MARITATI	130.54	2.611	1.373	1.9	1.29
	15KEN193	NGUSHISHI-MARITATI	130.54	2.611	1.373	1.9	1.29
	15KEN194	NGUSHISHI-MARITATI	136.92	2.738	1.481	1.85	1.46
	15KEN196	NGUSHISHI-MARITATI	196.38	3.928	2.254	1.74	1.28
	15KEN197	NGUSHISHI-MARITATI	139.63	2.793	1.492	1.87	1.57
	15KEN198	NGUSHISHI-MARITATI	225.05	4.501	2.327	1.93	1.69
	15KEN199	NGUSHISHI-MARITATI	154.14	3.083	1.564	1.97	1.71
	15KEN206	NGUSHISHI-MARITATI	52.62	1.052	0.563	1.87	1.34
	15KEN207	NGUSHISHI-MARITATI	83.11	1.662	0.911	1.82	1.43
	15KEN208	MAGUTU-LAIKIPIA	89.16	1.783	0.927	1.92	1.56
	15KEN209	MAGUTU-LAIKIPIA	32.23	0.645	0.369	1.75	0.86
	15KEN210	MAGUTU-LAIKIPIA	32.29	0.646	0.372	1.73	0.85
	15KEN211	MAGUTU-LAIKIPIA	33.74	0.675	0.37	1.82	1.27
	15KEN212	MAGUTU-LAIKIPIA	178.71	3.574	1.841	1.94	1.52
	15KEN213	MAGUTU-LAIKIPIA	56.66	1.133	0.62	1.83	1.12
	15KEN215	MAGUTU-LAIKIPIA	49.5	0.99	0.516	1.92	1.37
	15KEN216	MAGUTU-LAIKIPIA	53.74	1.075	0.434	2.48	1.19
	15KEN218	MAGUTU-LAIKIPIA	69.05	1.381	0.766	1.8	0.97
	15KEN219	MAGUTU-LAIKIPIA	68.82	1.376	0.74	1.86	1.11
	15KEN226	MAGUTU-LAIKIPIA	137.27	2.745	1.605	1.71	1.2
	15KEN227	MAGUTU-LAIKIPIA	75.92	1.518	0.853	1.78	0.91
	15KEN228	MAGUTU-LAIKIPIA	45.03	0.901	0.473	1.9	1.16
	15KEN229	MAGUTU-LAIKIPIA	62.41	1.248	0.708	1.76	1.28
	15KEN230	MAGUTU-LAIKIPIA	51.78	1.036	0.598	1.73	0.84
	15KEN231	MAGUTU-LAIKIPIA	68.45	1.369	0.808	1.69	1.05
	15KEN232	MAGUTU-LAIKIPIA	92.19	1.844	1.004	1.84	1.68
	15KEN233	MAGUTU-LAIKIPIA	51.99	1.04	0.608	1.71	0.83
	15KEN234	MAGUTU-LAIKIPIA	108.59	2.172	1.117	1.95	1.27
	15KEN235	MAGUTU-LAIKIPIA	35.29	0.706	0.375	1.88	1.29
	15KEN236	MAGUTU-LAIKIPIA	47.95	0.959	0.404	2.37	2.64
	15KEN237	MAGUTU-LAIKIPIA	102.72	2.054	1.104	1.86	1.35
	15KEN238	MAGUTU-LAIKIPIA	35.38	0.708	0.377	1.88	0.95
	15KEN249	MAGUTU-LAIKIPIA	82.21	1.644	0.915	1.8	1.04
NORTH RIFT	15KEN264	MEREWET	82.89	1.658	1.04	1.59	0.66
	15KEN291	MEREWET	102.26	2.045	1.057	1.94	1.37
	15KEN310	MEREWET	121.37	2.427	1.38	1.76	1.32

	15KEN311	RAY FARM (SERGOIT)	156.8	3.136	1.768	1.77	1.18
	15KEN312	RAY FARM (SERGOIT)	226.34	4.527	2.438	1.86	1.65
	15KEN313	RAY FARM (SERGOIT)	282.87	5.657	3.045	1.86	1.54
	15KEN314	RAY FARM (SERGOIT)	93.6	1.872	1.08	1.73	1.25
	15KEN315	RAY FARM (SERGOIT)	226.15	4.523	2.429	1.86	1.66
	15KEN316	RAY FARM (SERGOIT)	163.22	3.264	1.818	1.8	1.47
	15KEN317	RAY FARM (SERGOIT)	180.08	3.602	1.954	1.84	1.54
	15KEN318	RAY FARM (SERGOIT)	290.87	5.817	3.09	1.88	1.77
	15KEN320	RAY FARM (SERGOIT)	291.66	5.833	3.134	1.86	1.48
	15KEN321	RAY FARM (SERGOIT)	305.48	6.11	3.554	1.72	1.66
	15KEN322	RAY FARM (SERGOIT)	238.36	4.767	2.856	1.67	1.39
	15KEN323	RAY FARM (SERGOIT)	127.75	2.555	1.424	1.79	1.29
	15KEN324	RAY FARM (SERGOIT)	46.83	0.937	0.468	2	4.35
	15KEN325	RAY FARM (SERGOIT)	237.81	4.756	2.562	1.86	1.4
	15KEN326	RAY FARM (SERGOIT)	182.18	3.644	1.932	1.89	1.6
	15KEN327	RAY FARM (SERGOIT)	261.23	5.225	2.805	1.86	1.39
	15KEN328	RAY FARM (SERGOIT)	46.97	0.939	0.561	1.68	0.78
	15KEN329	RAY FARM (SERGOIT)	88.37	1.767	1.006	1.76	1.07
	15KEN330	RAY FARM (SERGOIT)	116.79	2.336	1.301	1.8	1.25
	15KEN331	RAY FARM (SERGOIT)	34.01	0.68	0.426	1.6	0.8
	15KEN331	RAY FARM (SERGOIT)	24.78	0.496	0.247	2.01	1.1
	15KEN332	RAY FARM (SERGOIT)	48.56	0.971	0.578	1.68	0.71
	15KEN333	RAY FARM (SERGOIT)	73.8	1.476	0.854	1.73	1.03
	15KEN337	RAY FARM (SERGOIT)	70.94	1.419	0.789	1.8	1.24
	15KEN340	RAY FARM (SERGOIT)	158.35	3.167	1.69	1.87	1.23
	15KEN343	RAY FARM (SERGOIT)	70.2	1.404	0.785	1.79	1.07
	15KEN344	RAY FARM (SERGOIT)	98.34	1.967	1.091	1.8	1.26
	15KEN345	RAY FARM (SERGOIT)	42.94	0.859	0.469	1.83	1.24
	15KEN346	RAY FARM (SERGOIT)	61.49	1.23	0.683	1.8	1
	15KEN347	RAY FARM (SERGOIT)	87.4	1.748	0.978	1.79	1.08
	15KEN348	RAY FARM (SERGOIT)	113.38	2.268	1.232	1.84	1.21
	15KEN349	RAY FARM (SERGOIT)	200.14	4.003	2.153	1.86	1.32
	15KEN350	RAY FARM (SERGOIT)	63.28	1.266	0.703	1.8	1.04
	15KEN354	RAY FARM (SERGOIT)	142.9	2.858	1.574	1.82	1.08
	15KEN355	RAY FARM (SERGOIT)	94.75	1.895	1.043	1.82	1.28
	15KEN360	RAY FARM (SERGOIT)	134.19	2.684	1.475	1.82	1.22
	15KEN361	RAY FARM (SERGOIT)	224.13	4.483	2.425	1.85	1.19
	15KEN362	RAY FARM (SERGOIT)	81.41	1.628	0.902	1.8	1.08
	15KEN363	RAY FARM (SERGOIT)	73.11	1.462	0.803	1.82	1.36
	15KEN364	RAY FARM (SERGOIT)	164.41	3.288	1.718	1.91	1.44
CENTRAL RIFT	15KEN365	KWA NDUSU RONGAI	145.29	2.906	1.582	1.84	1.17

15KEN369	KWA NDUSU RONGAI	79.06	1.581	0.78	2.03	1.53
15KEN377	KWA NDUSU RONGAI	104.03	2.081	1.063	1.96	1.52
15KEN378	KWA NDUSU RONGAI	120.17	2.403	1.37	1.75	0.85
15KEN379	KWA NDUSU RONGAI	134.05	2.681	1.411	1.9	1.67
15KEN380	KWA NDUSU RONGAI	88.89	1.778	0.977	1.82	1.04
15KEN381	KWA NDUSU RONGAI	102.18	2.044	1.062	1.92	1.39
15KEN382	KWA NDUSU RONGAI	82.41	1.648	0.881	1.87	0.74
15KEN383	KWA NDUSU RONGAI	132.92	2.658	1.372	1.94	1.44
15KEN384	KWA NDUSU RONGAI	156.22	3.124	1.567	1.99	1.56
15KEN385	KWA NDUSU RONGAI	142.14	2.843	1.46	1.95	1.46
15KEN386	KWA NDUSU RONGAI	134.26	2.685	1.378	1.95	1.57
15KEN387	KWA NDUSU RONGAI	191.25	3.825	1.98	1.93	1.47
15KEN388	KWA NDUSU RONGAI	161.18	3.224	1.628	1.98	1.82
15KEN389	KWA NDUSU RONGAI	106	2.12	1.078	1.97	1.7
15KEN390	KWA NDUSU RONGAI	130.25	2.605	1.32	1.97	1.65
15KEN391	KWA NDUSU RONGAI	129.07	2.581	1.339	1.93	1.98
15KEN392	KWA NDUSU RONGAI	219.12	4.382	2.322	1.89	1.69
15KEN393	KWA NDUSU RONGAI	220.14	4.403	2.403	1.83	1.16
15KEN394	KWA NDUSU RONGAI	300.61	6.012	3.096	1.94	1.81
15KEN395	KWA NDUSU RONGAI	236.08	4.722	2.412	1.96	1.81
15KEN396	KWA NDUSU RONGAI	252.21	5.044	2.6	1.94	1.66
15KEN397	KWA NDUSU RONGAI	149.02	2.98	1.54	1.93	1.57
15KEN398	KWA NDUSU RONGAI	53.76	1.075	0.599	1.79	1.17
15KEN399	KWA NDUSU RONGAI	160.4	3.208	1.628	1.97	1.91
15KEN406	KWA NDUSU RONGAI	102.42	2.048	1.069	1.92	1.69
15KEN407	KWA NDUSU RONGAI	244.23	4.885	2.458	1.99	2.06
15KEN408	KWA NDUSU RONGAI	309.78	6.196	3.032	2.04	2.23
15KEN409	KWA NDUSU RONGAI	128.11	2.562	1.257	2.04	2.05
15KEN410	KWA NDUSU RONGAI	150.61	3.012	1.504	2	2.14
15KEN411	KWA NDUSU RONGAI	98.33	1.967	0.961	2.05	1.88
15KEN412	KWA NDUSU RONGAI	110.92	2.218	1.096	2.02	1.96
15KEN413	KWA NDUSU RONGAI	150.02	3	1.518	1.98	1.99
15KEN414	KWA NDUSU RONGAI	105.98	2.12	1.05	2.02	2.09
15KEN415	KWA NDUSU RONGAI	100.18	2.004	0.98	2.05	1.78
15KEN416	MENENGAI	92.32	1.846	0.933	1.98	1.55
15KEN417	MENENGAI	238.81	4.776	2.469	1.93	1.66
15KEN418	MENENGAI	155.38	3.108	1.67	1.86	1.49
15KEN419	MENENGAI	67.98	1.36	0.705	1.93	1.52
15KEN420	MENENGAI	200.02	4	2.212	1.81	1.85
15KEN421	MENENGAI	109.83	2.197	1.135	1.94	1.34
15KEN422	MENENGAI	54.33	1.087	0.554	1.96	1.56

15KEN423	MENENGAI	229.72	4.594	2.208	2.08	2.22
15KEN424	MENENGAI	119.2	2.384	1.214	1.96	1.95
15KEN425	MENENGAI	218.6	4.372	2.275	1.92	1.47
15KEN426	MENENGAI	193.69	3.874	1.959	1.98	1.69
15KEN427	MENENGAI	135.77	2.715	1.423	1.91	1.42
15KEN428	MENENGAI	121.62	2.432	1.413	1.72	1.16
15KEN429	MENENGAI	345.64	6.913	3.61	1.91	1.98
15KEN430	MENENGAI	242.18	4.844	2.433	1.99	2.14
15KEN431	MENENGAI	472.88	9.458	4.82	1.96	2.2
15KEN432	MENENGAI	156	3.12	1.593	1.96	1.93
15KEN433	MENENGAI	85.1	1.702	0.852	2	1.63
15KEN434	MENENGAI	655.66	13.113	6.617	1.98	2.24
15KEN435	MENENGAI	193.31	3.866	1.946	1.99	1.98
15KEN436	MENENGAI	461.03	9.221	5.447	1.69	1.08
15KEN437	MENENGAI	283.86	5.677	2.855	1.99	2.2
15KEN438	MENENGAI	229.12	4.582	2.307	1.99	1.97
15KEN439	MENENGAI	409.73	8.195	4.315	1.9	1.66
15KEN440	MENENGAI	1942.58	38.852	21.798	1.78	1.52
15KEN441	MENENGAI	347.05	6.941	3.539	1.96	1.9
15KEN442	MENENGAI	443.2	8.864	4.512	1.96	2
15KEN443	MENENGAI	311.84	6.237	3.258	1.91	1.73
15KEN444	MENENGAI	277.94	5.559	2.768	2.01	2.14
15KEN445	MENENGAI	603.68	12.074	6.812	1.77	1.83
15KEN446	MENENGAI	333.16	6.663	4.22	1.58	1.57
15KEN447	MENENGAI	572.08	11.442	6.294	1.82	1.52
15KEN448	MENENGAI	244.81	4.896	2.402	2.04	1.87
15KEN450	MENENGAI	46.11	0.922	0.425	2.17	1.74
15KEN485	PIAVE	138.91	2.778	1.701	1.63	1.03
15KEN486	PIAVE	139.12	2.782	1.434	1.94	1.44
15KEN487	PIAVE	110.03	2.201	1.167	1.89	1.49
15KEN488	PIAVE	132.87	2.657	1.493	1.78	1.16
15KEN490	PIAVE	50.16	1.003	0.541	1.85	1.5
15KEN497	PIAVE	82.69	1.654	0.862	1.92	1.52
15KEN521	PIAVE	146.78	2.936	1.524	1.93	1.7

Appendix II: Information on 104 samples genotyped, their region and the SSR-MLGs.

Sample	Region	MLG	PSex	PValue
KEN162	MT	MLG_1	2.33E-10	0
KEN191	MT	MLG_1	2.33E-10	0
KEN206	MT	MLG_1	2.33E-10	0
KEN182	MT	MLG_2	2.33E-15	0
KEN198	MT	MLG_2	2.33E-15	0
KEN212	MT	MLG_2	2.33E-15	0
KEN161	MT	MLG_2	2.33E-15	0
KEN190	MT	MLG_2	2.33E-15	0
KEN213	MT	MLG_2	2.33E-15	0
KEN229	MT	MLG_2	2.33E-15	0
KEN237	MT	MLG_2	2.33E-15	0
KEN174	MT	MLG_2	2.33E-15	0
KEN215	MT	MLG_2	2.33E-15	0
KEN230	MT	MLG_2	2.33E-15	0
KEN238	MT	MLG_2	2.33E-15	0
KEN164	MT	MLG_2	2.33E-15	0
KEN175	MT	MLG_2	2.33E-15	0
KEN192	MT	MLG_2	2.33E-15	0
KEN216	MT	MLG_2	2.33E-15	0
KEN249	MT	MLG_2	2.33E-15	0
KEN165	MT	MLG_2	2.33E-15	0
KEN176	MT	MLG_2	2.33E-15	0
KEN193	MT	MLG_2	2.33E-15	0
KEN208	MT	MLG_2	2.33E-15	0
KEN218	MT	MLG_2	2.33E-15	0
KEN232	MT	MLG_2	2.33E-15	0
KEN178	MT	MLG_2	2.33E-15	0
KEN194	MT	MLG_2	2.33E-15	0
KEN209	MT	MLG_2	2.33E-15	0
KEN233	MT	MLG_2	2.33E-15	0
KEN179	MT	MLG_2	2.33E-15	0
KEN196	MT	MLG_2	2.33E-15	0
KEN210	MT	MLG_2	2.33E-15	0

KEN226	MT	MLG_2	2.33E-15	0
KEN211	MT	MLG_2	2.33E-15	0
KEN227	MT	MLG_2	2.33E-15	0
KEN422	CR	MLG_2	2.33E-15	0
KEN430	CR	MLG_2	2.33E-15	0
KEN423	CR	MLG_2	2.33E-15	0
KEN431	CR	MLG_2	2.33E-15	0
KEN439	CR	MLG_2	2.33E-15	0
KEN447	CR	MLG_2	2.33E-15	0
KEN416	CR	MLG_2	2.33E-15	0
KEN424	CR	MLG_2	2.33E-15	0
KEN448	CR	MLG_2	2.33E-15	0
KEN425	CR	MLG_2	2.33E-15	0
KEN441	CR	MLG_2	2.33E-15	0
KEN450	CR	MLG_2	2.33E-15	0
KEN314	NR	MLG_2	2.33E-15	0
KEN333	NR	MLG_2	2.33E-15	0
KEN326	NR	MLG_2	2.33E-15	0
KEN318	NR	MLG_2	2.33E-15	0
KEN355	NR	MLG_2	2.33E-15	0
KEN360	NR	MLG_2	2.33E-15	0
KEN361	NR	MLG_2	2.33E-15	0
KEN169	MT	MLG_3	2.11E-15	0
KEN429	CR	MLG_3	2.11E-15	0
KEN437	CR	MLG_3	2.11E-15	0
KEN417	CR	MLG_3	2.11E-15	0
KEN418	CR	MLG_3	2.11E-15	0
KEN199	MT	MLG_4	0	0
KEN207	MT	MLG_4	0	0
KEN170	MT	MLG_4	0	0
KEN197	MT	MLG_4	0	0
KEN365	CR	MLG_4	0	0
KEN383	CR	MLG_4	0	0
KEN399	CR	MLG_4	0	0
KEN384	CR	MLG_4	0	0
KEN392	CR	MLG_4	0	0
KEN406	CR	MLG_4	0	0
KEN414	CR	MLG_4	0	0
KEN446	CR	MLG_4	0	0
KEN377	CR	MLG_4	0	0
KEN385	CR	MLG_4	0	0
KEN393	CR	MLG_4	0	0
KEN407	CR	MLG_4	0	0

KEN378	CR	MLG_4	0	0
KEN394	CR	MLG_4	0	0
KEN408	CR	MLG_4	0	0
KEN395	CR	MLG_4	0	0
KEN380	CR	MLG_4	0	0
KEN396	CR	MLG_4	0	0
KEN410	CR	MLG_4	0	0
KEN426	CR	MLG_4	0	0
KEN485	CR	MLG_4	0	0
KEN323	NR	MLG_4	0	0
KEN315	NR	MLG_4	0	0
KEN324	NR	MLG_4	0	0
KEN291	NR	MLG_4	0	0
KEN317	NR	MLG_4	0	0
KEN310	NR	MLG_4	0	0
KEN327	NR	MLG_4	0	0
KEN320	NR	MLG_4	0	0
KEN091	SR	MLG_4	0	0
KEN049	SR	MLG_4	0	0
KEN119	SR	MLG_4	0	0
KEN055	SR	MLG_4	0	0
KEN121	SR	MLG_4	0	0
KEN067	SR	MLG_4	0	0
KEN130	SR	MLG_4	0	0
KEN264	NR	MLG_5	1.11E-15	0
KEN325	NR	MLG_5	1.11E-15	0
KEN350	NR	MLG_5	1.11E-15	0
KEN123	SR	MLG_5	1.11E-15	0

Appendix III: Analysis table from the RClone results showing genotypic diversity, richness and evenness indices calculation.

	N	Lineage	nb_L	nb_all	SE	Fis	pval_2sides
1	41	MLG	4	3	0.233549683248457	-0.528790316304257	NA
2	37	MLG	3	3	0.233549683248457	-0.413616176369373	NA
3	18	MLG	3	2.90909090909091	0.211253637065859	-0.389147995963676	NA
4	8	MLG	2	2	0	-0.704545454545455	NA

	Fis_WR	pval_2sides.1	R	Pareto_index	Sp_Loiselle
1	-0.631149455679109	NA	0.075	0.06449106568624	NA
2	-0.120385506311588	NA	0.0555555555555556	0.308640082951389	NA
3	-0.208892404847529	NA	0.117647058823529	0.608918485826371	NA
4	-0.363636363636364	NA	0.142857142857143	0.0686215613240665	NA

Appendix III Continued

	H..	J.	D	V	Hill
1	0.68367538770729	0.493167545711573	0.345121951219512	0.323350377404431	1.52700186219739
2	0.927160387397537	0.843937753983773	0.576576576576577	0.812283737024222	2.36170212765957
3	1.02633063329449	0.934206401913405	0.660130718954248	0.906666666666667	2.94230769230769
4	0.376770161256437	0.543564443199596	0.25	0	1.33333333333333

1= Mount Kenya. 2= Central Rift. 3= North Rift. 4= South Rift

H", the Shannon-Wiener index estimator (Pielou 1966)

N, the number of units in KE2015kkk

J', the Pielou evenness index(Pielou 1975)

Lineage, MLG or MLL

D', the Simpson complement unbiased (Pielou 1969 ; Gini 1912 ; Peet 1974)

nb_L, the number of MLG/MLL

V, the Simpson complement index (Hurlbert 1971 ; Fager 1972)

nb_all, the mean number of alleles

Hill, the reciprocal of Simpson index unbiased (Hurlbert 1971 ; Hill 1973)

SE, the standard error of nb_all,

Fis, on ramets if diploid data

pval_2sides, the two-sided p-value of Fis if nbrepeat,

Appendix III Continued

Fis_WR, on genets if diploid data

pval_2sides, the two-sided p-value of Fis_WR if nbrepeat,

R, the clonal diversity index (Dorken & Eckert 2001; Ellstrand & Roose 1987),

Pareto_index, the index of Pareto

Sp_Loiselle, Sp index computed on ramets with Loiselle kinship results used to quantify Spatial Genetic Structure (Vekemans and Hardy, 2004)

Appendix IV: Analysis of Molecular Variance (AMOVA) output to partition diversity among and within populations

Results of Analysis of Molecular Variance

Input as Allelic Distance Matrix for F-Statistics Analysis

No. Samples 104

No. Pops 4

No. Regions 1 (Kenya)

No. Permutations 999

Pop	MT	CR	NR	SR
n	82	74	36	16
SSWP	173.341	217.635	107.778	37.750

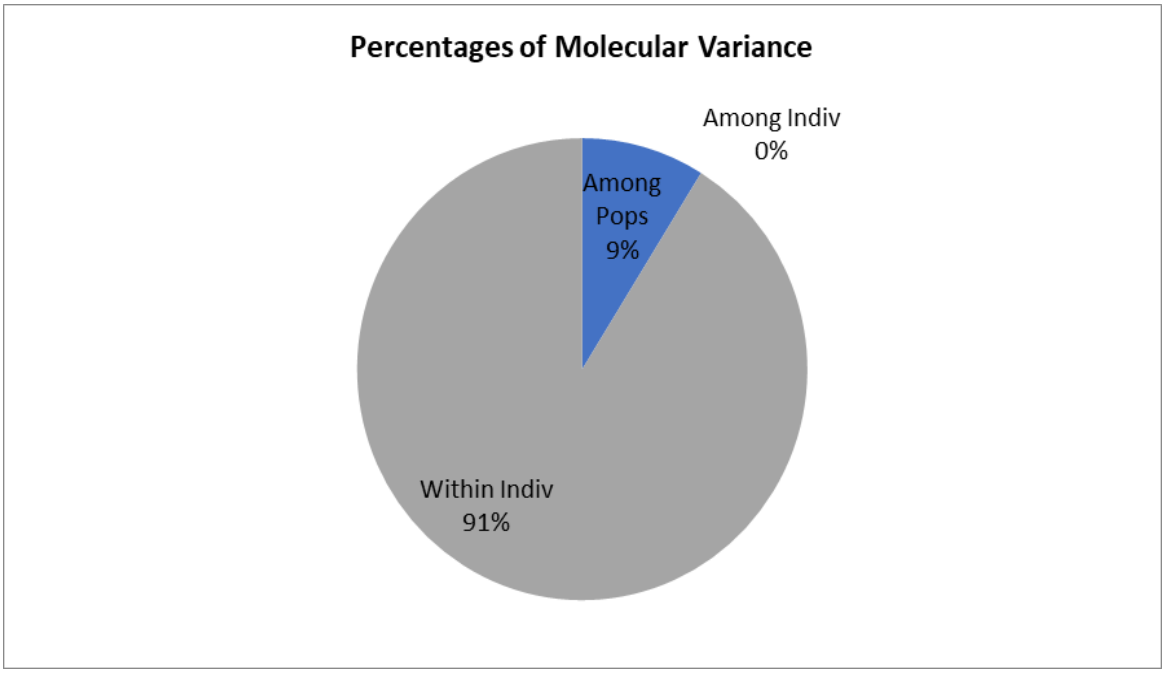
Summary AMOVA Table

Source	df	SS	MS	Est. Var.
Among Pops 9%	3	57.750	19.250	0.380
Among Indiv 0%	100	130.00	1.300	0.000
Within Indiv 91%	104	406.500	3.909	3.909
Total 100%	207	594.255		4.288

F-Statistics	Value	P(rand >= data)
Fst	0.127	0.001
Fis	-0.501	1.000
Fit	-0.310	1.000
Fst max	0.476	
F'st	0.267	

Nm	1.715

Appendix IV Continued



Probability, P(rand >= data), for Fst, Fis and Fit is based on standard permutation across the full data set.

Fst = AP / (WI + AI + AP) = AP / TOT

Fis = AI / (WI + AI)

Fit = (AI + AP) / (WI + AI + AP) = (AI + AP) / TOT

Nm = [(1 / Fst) - 1] / 4

Key: AP = Est. Var. Among Pops, AI = Est. Var. Among Individuals, WI = Est. Var. Within Individuals

Warning! Negative Among Individuals. Est. Var. of -1.304 converted to zero for pie chart presentation!