CHAPTER I

1.1 Introduction

Malaria is a mosquito-borne infectious disease and is caused by protozoan parasites of the genus *Plasmodium*. Five species that can cause the human disease are *Plasmodium falciparum, P. vivax, P. malariae*, and *P. ovale*. *P. knowlesi*, where the natural host is the long- and the pig-tailed monkeys, has been reported as the fifth parasite species (White, 2008). *Plasmodium* parasite undergoes different stages in the life cycle both occuring in mosquitoes and humans. The parasite is transmitted to humans by female mosquitoes of the *Anopheles* genus. About 60 of the 390 species of *Anopheles* mosquito have been implicated as vectors of malaria parasite (Irion, 2000). *P. falciparum* causes the most severe form of malaria and is responsible for mortality and morbidity.

Malaria is globally distributed in around 108 countries, affecting mainly tropical areas of Asia, Africa, and Latin America. It causes devastating toll in lives, medical costs and days of labor lost (Sachs and Malaney, 2002). It has been reported that about 300 to 400 million people are infected with malaria, causing an annual mortality of about 1 to 2 million people (WHO, 2004). However, epidemiological survey indicates that these numbers might be highly underestimated (Snow et al., 2005). Aregawi et al. (2008) reported that there were 3.3 billion of malaria cases worldwide in 2008. In 2009, WHO reported 243 million of malaria cases with mortality of 863,000 (WHO, 2009).

Information on the nature and extent of genetic diversity within *P. falciparum* is essential for understanding the mechanism underlying the pathology of malaria, the acquisition of immunity, the spread of drug resistance and the condition of the transmission. The Merozoite surface proteins *MSP-1* and *MSP-2* are highly polymorphic markers and allele polymorphisms have been reported in block 2 of the *MSP-1* gene and the central repetitive domain (block 3) of the *MSP-2* gene. Families differing in nucleotide sequences and in number of repetitive sequences (length variation) have been used for genotyping purposes. Studies on genetic diversity, the differentiation of different strains within a *Plasmodium* species, and presence of multiple parasite strains in individual host have been reported from different regions around the globe (Aubouy et al., 2003; Zakeri et al., 2005).

1.2 Justification of the study

In Yemen, malaria is one of the most serious health problems. About 60% of the population lives in areas with malaria transmission. *P. falciparum* is the predominant species accounting for more than 90% of malaria cases. Genetic diversity provides unique insights into the transmission dynamic in such regions of malaria parasite populations. Also, genetic polymorphism of *P. falciparum* is a major consideration in the vaccination programs against malaria infection. The polymorphic regions of *MSP-1* and *MSP-2* genes are essential genetic markers which can provide information on the complexity of *P. falciparum* infection and level of transmission. Moreover, the ability of *MSP-1* and *MSP-2* in inducing strong immune response in human makes them as

potential vaccine candidates against blood stages of the parasite. Until now, no study has specifically estimated the genetic complexity of *P. falciparum* in Yemen.

1.3 General Objective

This study aimed to genotyping of Yemen isolates *P. falciparum*, based on *MSP-1* and *MSP-2* genes.

1.4 Specific Objectives

1. To determine prevalence and distribution of allelic families of *MSP-1* and *MSP-2* genes in different malaria endemic areas of Yemen (highland, hinterland and coastland).
2. To compare the genetic complexity of *P. falciparum* between these malaria endemic areas.
3. To correlate between the complexity of *P. falciparum* and age of patients as well as parasitaemia.

CHAPTER II

LITERATURE REVIEW

2.1 *Plasmodium* species and life cycle

Malaria is due to the protozoa of the genus *Plasmodium*. The four species of *Plasmodium* that infect humans (*P. falciparum, P. malariae*, *P. vivax* and *P. ovale*) have a common life cycle with minor differences. *P. falciparum* causes the most serious type of disease, which is caused by the ability of the parasite to penetrate the brain causing rapidly developing encephalopathy (cerebral malaria).

P. falciparum is a parasite that lives in the mosquito vector and human host. It has a complex life cycle and consists of distinct stages in the *Anopheles* mosquito vector and human host (Figure 2.1).

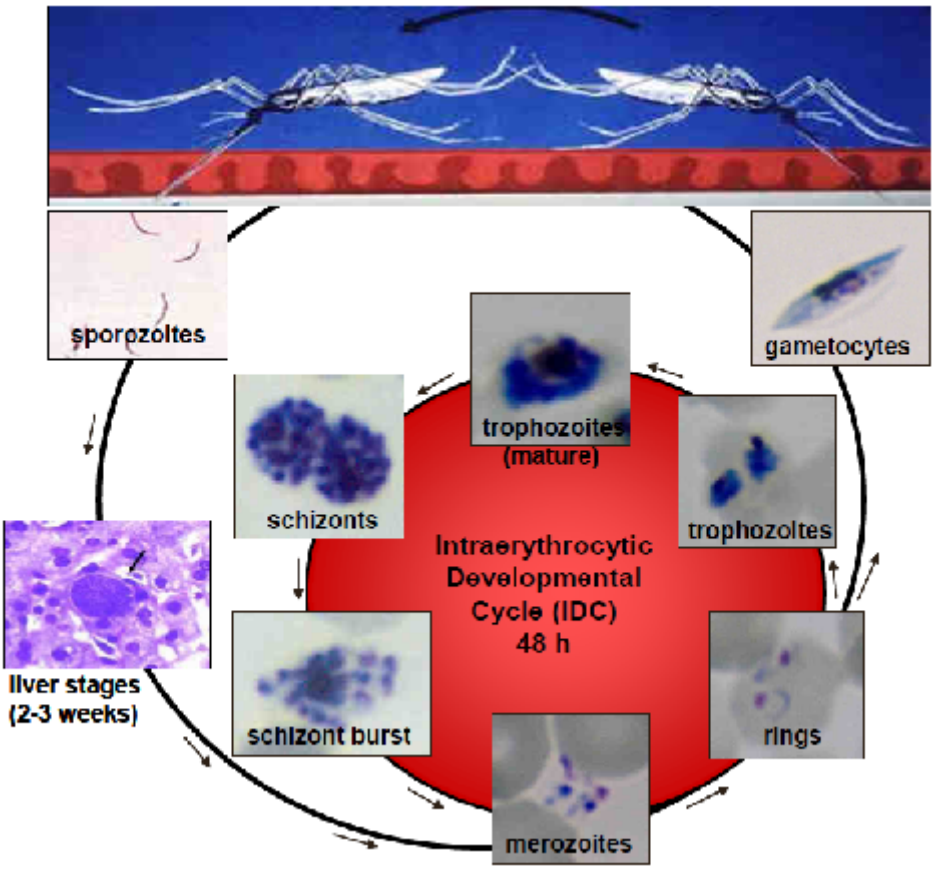

Figure 2.1: Giemsa-stained thin smears depicting the life cycle of 3D7 *P. falciparum* (adapted from Anna Catharina, 2008).

When an infected female *Anopheles* mosquito bites the human skin, sporozoites are passed along with the anticoagulant saliva into the blood. The sporozoites readily travel through the bloodstream to the liver where they will settle in liver cells. In the liver cells, the sporozoites undergo an initial growth followed by asexual replication (liver schizogony), forming a mature liver schizont containing up to 30,000 merozoites. The liver schizogony takes 5-7 days and during this time, the patient is asymptomatic. As the mature liver schizonts rupture, merozoites are released. Majority are engulfed by liver macrophages i.e. Kupffer cells however, merozoites that escape rapidly infect erythrocytes (red blood cells). In the erythrocyte, the merozoite transforms to an immature trophozoite (ring form), then to a mature trophozoite, followed by asexual reproduction to a mature schizont containing around 10 to 20 merozoites. The merozoites are released upon erythrocyte rupture and instantly invade new erythrocytes. The duration of the erythrocytic cycle lasts for 48 hours (Garcia et al., 2006).

A characteristic of *P. falciparum* is sequestration i.e. binding of infected erythrocytes to endothelium in the deep vascular system during the second half of the erythrocytic cycle. Therefore, only ring forms and early trophozoites are detectable in peripheral blood (Aikawa et al., 1990). Parasite debris is released due to erythrocyte rupture and this causes host responses such as fever, cytokines release and the symptomatic phase of the infection starts. The clinical manifestations vary from asymptomatic infections to severe life-threatening conditions. Some merozoites do not undergo further asexual replication, instead they develop into male and female gametocytes. Gametocytes circulating in the blood are extracted by mosquito during its feeding on the human blood. Within the mosquito's gut, the gametocytes are triggered by the presence of specific mosquito

factors and the drop in temperature to form male or female gametes (Liljander, 2010). A male and a female gamete fuse to form a diploid zygote which undergoes meiosis and produces ookinete. The ookinete penetrates the mosquito's mid-gut wall, transforms into an oocyst that produces a large number of haploid sporozoites through repeated mitotic divisions. These will migrate and invade the salivary glands from which they can be injected into the human host when the mosquito takes a blood meal, thus starting the life cycle of the parasite again. The process takes 10-18 days depending on the *Plasmodium* species. The mosquito remains infectious for 1-2 months (Liljander, 2010).

2.2 The mosquito vector

Human malaria is transmitted by female mosquitoes of the genus *Anopheles*. Nearly 60 of the 390 species of *Anopheles* mosquito transmit the malaria parasite (Irion, 2000), depending on the region and the environment (Subbarao and Sharma, 1997). Anophelines that can transmit malaria are found in malaria endemic areas and in the areas where malaria has been eliminated. Thus, the latter areas are constantly at risk of re-introduction of the disease. *Anopheles* mosquitoes are also able to transmit the Timorese filaria, *Brugia timori*; *W. bancrofti* (filarial worm), several arboviruses including Venezuelan equine encephalitis, and western and eastern equine encephalitis.

2.3 Distribution of Malaria Worldwide

According to the World Malaria Report (2009) ((Figure 2.2) malaria is prevalent in 108 countries of the tropical and semitropical world, with 35 countries in central Africa bearing the highest burden of cases and deaths. Of the 35 countries that account globally for ~98% of malaria deaths, 30 are located in sub-Saharan Africa, accounting for 98.5% of the deaths in Africa, with four countries alone accounting for ~50% of deaths on the continent (Nigeria, Democratic Republic of Congo, Uganda and Ethiopia). In sub-Saharan Africa, approximately 365 million cases occurred in 2002 and 963 thousand deaths in 2000, equating to 71% of worldwide cases and 85.7% of worldwide deaths. Almost 1 out of 5 deaths of children under 5 in Africa is due to malaria. With increased efforts in controlling malaria in Africa in the recent years, it is reported that a total of 11 countries and one area in the African Region showed a reduction of more than 50% in either confirmed malaria cases or malaria admissions and deaths (Algeria, Botswana, Cape Verde, Eritrea, Madagascar, Namibia, Rwanda, Sao Tome and Principe, South Africa, Swaziland, Zambia, and Zanzibar, United Republic of Tanzania), whereas there was evidence of an increase in malaria cases in 3 countries in 2009 (Rwanda, Sao Tome and Principe, and Zambia). (WHO, 2009)

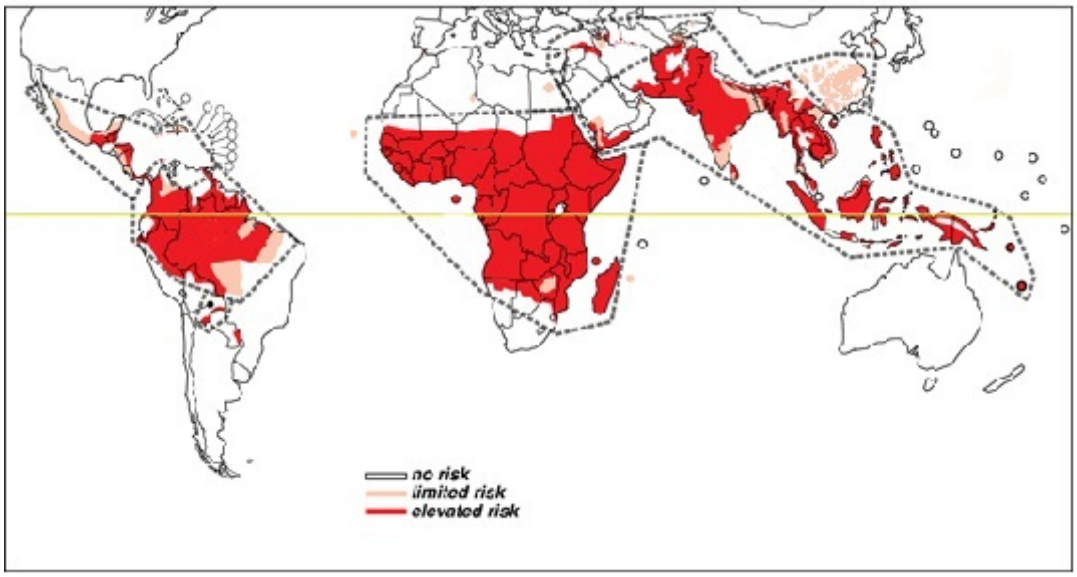

Figure 2.2: Distribution of malaria worldwide (WHO, 2009).

In Yemen fifteen *Anopheles* species are documented. *Anopheles* a*rabiensis* is the main vector but *A. culicifacies* plays an important role in the transmission of malaria in the coastal areas while *A. sergenti* has been reported to be a vector in the mountainous hinterland and highland areas (NMCP, 2002).

2.4Transmission and epidemiology of malaria

Transmission of malaria is constrained to the geographical area in which the *Anopheles* mosquitoes flourish depending on the favorable climate especially the temperature in tropical and sub-tropical regions. Native malaria can be either endemic or epidemic. Endemic transmission is characterized by firm transmission over long period of time. The transmission may be stable which is characterized by continuous transmission i.e., constant over many years with or without seasonal instability, or unstable transmission with remarked fluctuations. Entomological inoculation rate (EIR) is calculated as the total mosquito biting rate, while the sporozoite rate is the number of infective mosquito bites received per person per year which is used to determine the level of transmission of malaria. In stable transmission, the EIR is greater than 10 per year, while in unstable transmission the EIR is between 1-5 per year (WHO, 2010). Malaria epidemics might take place in area with low and unstable transmission and which is characterized by a sudden increase in the number of clinical cases.

2.5 Genetic diversity of *P. falciparum* infections based on genes encoding merozoite surface proteins (*MSP*)

The introduction of PCR-based genotyping techniques in malaria research has substantially improved the understanding of the parasite biology and epidemiology. A number of highly polymorphic genetic markers for *P. falciparum* have been characterized and can be used to distinguish individual parasite populations. The most widely used markers for genotyping of *P. falciparum* are the genes encoding merozoite surface protein 1 (*MSP-1*), merozoite surface protein 2 (*MSP-2*) and glutamate-rich protein (*glurp*). These genes are suitable to characterize parasite populations since they are unlinked single copy genes and remain stable throughout the erythrocytic life. In epidemiological studies, genotyping is used to investigate infection diversity and number of infecting parasite clones in relation to factors such as transmission intensity and host immunity. In antimalarials drug trials, genotyping is recommended to define treatment outcome by differentiating recrudescent parasites from new infections (WHO, 2008). In order to distinguish recurrence from a new infection, genotyping is often performed by stepwise adding several consecutive markers (*MSP-1*, *MSP-2* and glurp) (Mugittu et al., 2006). In epidemiological studies assessing infection diversity, a single marker is often sufficient and MSP-2 has been shown to be the most informative marker (Farnert et al., 2001). However, analysis of more than one marker gene is the best strategy for detecting genetic diversity, because the probability of different clones sharing the same genotype by chance is substantially reduced (Schoepflin et al., 2009).

Merozoite surface protein (*MSP*) is one of the proteins of the erythrocytic stage of the parasite's life cycle. Many merozoite surface proteins have been described for

Plasmodium. These proteins are involved in the initial recognition and attachment with the erythrocyte surface. Nine *MSP* (1-9) have been identified in either *P. vivax* or in *P. falciparum*. Some *MSP* have been found in other *Plasmodium* sp., such as *P. knowlesi* (*MSP-1*, *MSP-3*a, *MSP-3*b, *PkMSP-1*10). The *MSP-1* and *MSP-2* induce host's immune responses and they have been considered as potential targets for vaccine development (Kumar et al., 2002). *MSP-1* genes have been well characterized in *P. falciparum, P. vivax* and rodent malaria species (Gibson et al., 1992; Miller et al., 1993), and have been proposed to contain epidermal growth factor (EGF)-like domains (Blackman and Holder, 1992). *MSP-1* has been shown to be present in *P. cynomolgi, P. yoelii, P. berghei,* and *P. chabaudi* (Miller et al., 1981; Hadley et al., 1984; Holder and Freeman, 1984; Lewis 1989; Hendrix et al., 1990; O'dea et al., 1993; Toebe et al., 1997). The *P. falciparum* MSP are the most well characterized, especially *MSP-1* which is a leading vaccine candidate and is ubiquitous among other species of *Plasmodium*. *P. falciparum MSP-1* is a 200 kDa protein present on the surface of merozoites as a noncovalent complex of four fragments 30, 38, 42, and 83 kDa (Holder and Freeman, 1984). The 42 kDa C-terminal fragment binds the complex to the surface of the merozoite via a glycosyl phosphatidylinositol (GPI) anchor (Schofield et al., 1996). Further processing of this fragment yields two fragments of 33 and 19 kDa. The 33 kDa form is shed with the rest of the complex during erythrocyte invasion while the 19 kDa form is still detected after invasion within the early trophozoite (Blackman and Holder, 1992). Monoclonal antibodies against the C-terminus inhibit merozoite invasion (Chappel and Holder, 1993). The *MSP-1* gene has 7 variable blocks that are separated either by conserved or semi-conserved regions. Block 2, a region near the N-terminal of the *MSP-1* gene, is the most

polymorphic part of the antigen and appears to be under the strongest diversifying selection within natural populations (Holder and Blackman, 1994). Four different allelic types of block 2 have been identified: MAD20, K1, RO33, and MR (Happi et al., 2004).

Four other *P. falciparum MSP* (*MSP-2*, *MSP-4*, *MSP-5*, and *MSP-8*) are predicted to bind to the plasma membrane via GPI anchors. *MSP-2*, *MSP-4*, and *MSP-5* are closely linked within a 5 kb region on chromosome 2 (Marshall et al., 1998). *MSP-2* is synthesized early in schizogony as a 46-53 kDa protein and appears unprocessed on the merozoite surface after release from the erythrocyte (Donachie et al. 1989). *MSP-3*, synonymous with secreted polymorphic antigen associated with merozoites (SPAM), is soluble and found within the parasitophorous vacuole (Silva et al. 1994; Tayoun et al., 1994). The 40 kDa *MSP-4* and 40 kDa *MSP-5* both localize to the surface and possess an EGF-like domain at their C-termini, similar to *MSP-1* (Silva et al., 1997; Marshall et al., 1998). *MSP-8* has two EGF-like domains, localizes to the surface of merozoites, trophozoites, and schizonts (Wu et al., 2001). Two paralogues of *MSP-3* (H101 and H103) has been found in *P. falciparum* (Mills et al., 2005). Both *MSP-6* and the 22 kDa *MSP-7* are linked directly with the *MSP-1* complex on the surface of merozoites (Ling et al., 2001). *MSP-8* is the most highly conserved of the MSP molecules, and appears to be proteolytically processed into four fragments (98, 50, 25, and 19 kDa) similar to *MSP-1* (Wu et al., 2001). *MSP-9* was first identified in *P. vivax, P. knowlesi,* and *P. cynomolgi* and is related to the *P. falciparum* acidic-basic repeat antigen (ABRA) molecule (Barnwell et al., 2002). Functional studies indicated that *P. falciparum* ABRA/*MSP-9* has serine proteinase

activity and antibodies that inhibit parasite growth in vitro implicating its role in the invasion process (Haynes et al., 1992; Rao et al., 2001).

MSP-2 is much smaller (~28kDa) than *MSP-1*, and is highly polymorphic (Felger et al., 1997; Marshall *et al.,* 1994; Snewin *et al.,* 1991). A single centrally-located block of diverse sequence in *MSP-2* is flanked by highly conserved N-and C-terminal sequences. A very large number of *MSP-2* alleles have been identified but like *MSP-1*, *MSP-2* is dimorphic in that the sequences can be readily classified into one of two types, exemplified by the FC27 and the 3D7 alleles (Fenton *et al.,* 1989; Smythe *et al.,* 1990). The dimorphism in *MSP-2* is most clearly seen in the variable non-repetitive sequences which flank the central block of repeats. The repeats are highly polymorphic but different types of repeats are seen in the two families of *MSP-2* alleles. Alleles of the FC27-type contain one to three tandem copies of a sequence encoding a 32-residue repeat followed by one to five tandem copies of a sequence encoding a 12-residue repeat. Recently, FC27-type alleles have been described in which the NAP sequence N-terminal to the first 32-residue repeat is amplified to generate two to 23 tandem copies of this sequence (Beck and Felger, 1997). In contrast, the *MSP-2* alleles of the 3D7 type encode variable numbers of repeats rich in alanine, glycine and serine. The 32-residue repeat in the FC27 family exhibits diversity among different isolates with amino acid substitutions occurring in a restricted region close to the N-terminal end of the repeat. The sequence motif for this region is also common in the short repeats found in the 3D7 allelic family and it has been suggested that this provides a hot spot for intragenic recombination in *MSP-2* (Beck and Felger, 1997). Although added diversity in *MSP-2* is generated as a result of

intragenic recombination between the two allelic families (Marshall et al., 1991) the frequency of such alleles appears to be low.

Due to their polymorphic features, the *MSP-1* and *MSP-2* genes have been employed as polymorphic markers in studies of malaria transmission dynamics in natural isolates of *P. falciparum*. Various approaches have been used for genotyping of *P. falciparum* such as amplified fragment length polymorphism (AFLP) (Rubio et al., 2001), PCR-restriction fragment length polymorphism (PCR-RFLP) (Mayor et al., 2003), microsatellite analysis (MS) (Anderson et al., 2000), fluorogenic PCR (Decuypere et al., 2003), sequence analysis (Basco et al., 2004), matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF) (Marks et al., 2004), minisatellite variant repeat (MVR) mapping (MacLeod, 2004) and ligase detection reaction-fluorescent microsphere assay (LDR-FMA) (Dent et al., 2007).

The diversity of *P. falciparum* within a single host can reach up to about ten different genotypes at a given time – but the mean number in a population is about 2-4. The number of genotypes within an infection is a result of interaction between several factors i.e., intensity of malaria transmission in the area, individual exposure to infective mosquitoes, natural and acquired immunity as well as chemoprophylaxis or recent treatment. The parasite population counted in a single blood sample may be underestimated from the actual number considering an extensive dynamics of parasite population within the host (Daubersies et al., 1996; Farnert et al., 1997; Bruce et al., 2000).

Although different parasites may be found over time, the number of genotypes at a given

time is relatively constant (Farnert et al., 1999) suggesting that single sample is sufficient in epidemiological studies. Malaria transmission intensity can influence the genetic diversity of the parasite population. Sexual reproduction of the parasites in the mosquito allows intragenic recombination and chromosome assortments events which are more frequent in high transmission areas (Walliker, 1983). Studies comparing different transmission intensities indeed established that *P. falciparum* populations, both in an area and in an individual, are more diverse in areas of high transmission compared to areas with low or seasonal transmission (Babiker et al., 1997; Konate et al., 1999). Higher transmission levels are associated with a non-linear increase in the number of parasite genotypes per host (Arnot 1998; Babiker et al., 1999; Bendixen et al., 2001). However, even in areas with the most intense transmission individuals can be infected with only a limited number of different genotypes (Felger et al., 1994; Engelbrecht et al., 2000).

Many studies have been conducted on the epidemiology of genetic diversity of the *P. falciparum*, in different geographical areas with different transmission states, focusing on the polymorphism of *MSP-1* and *MSP-2* genes (Tables 2.1 and 2.2). A few studies have searched for allelic markers of parasite virulence and found that alleles of the FC27 family of *MSP-2* are associated with malaria morbidity (Felger et al., 1994; Beck et al., 1997; Genton et al., 2002). Extensive parasite diversity is found in highly endemic areas, thus infected human carry multiple genotypes (Babiker et al., 1995; Paul et al., 1998), while the low transmission areas are characterized by limited genetic diversity and most of the infections are monoclonal (Haddad et al., 1999; Hoffmann et al., 2001). In the

Eastern Mediterranean Region, studies have been carried out in Iran (Zakeri et al., 2005; Heidari et al., 2007), and Sudan (Babiker, 1995).

2.6 Status of malaria in Yemen

Yemen is located in the Middle East at the southern tip of Arabian Peninsula in Southwest Asia. It is bordered on the north by Saudi Arabia, on the east by Oman, on the south by the Gulf of Aden and on the west by the Red Sea (Figure 2.4). The geography of Yemen can be divided into four main regions: the coastal plains in the west, the western highlands, the eastern highlands, and the Rub al Khali in the east. The total population is estimated at 23,580,000 people according to Department of Economic and Social Affairs Population Division (2009), within an area of about 527,968 km^2.

Yemen belongs to the Eastern Mediterranean Region which contains five other countries, (Pakistan, Afghanistan, Sudan, Somalia Djibouti), with areas of high malaria transmission, and three countries (Iraq, Iran, and Saudi Arabia) with geographically limited malaria transmission. *P. falciparum* is the predominant species of parasite in Saudi Arabia, Djibouti, Yemen, and Sudan, while the most common species in Afghanistan, Pakistan, Iran and Iraq is *P. vivax*. The Eastern Mediterranean region had 890,000 confirmed cases in 2008. Four countries accounted for 90% of the estimated cases: Afghanistan (7%); Pakistan (18%); Somalia (10%) and Sudan (62%). Afghanistan, Iran, Iraq and Saudi Arabia have reported downward trends in malaria frequency. Djibouti, Pakistan, Somalia, Sudan have not reported decrease in the number of cases (WHO, 2008).

Figure 2.3: Map of Yemen

In Yemen, malaria causes serious health problems, where approximately 45% of the population has been estimated to be at risk (WHO, 2008). *P. falciparum* is distribuated and it accounts for more than 90% of malaria cases reported. Cases in Yemen decreased gradually to 900,000 in 2008, as compared to 3.2 million cases in 2001.The mortality and morbidity rates have decreased gradually in recent years, although malaria remains as one of the most serious problems threatening human health in Yemen. Information on the extent and nature of diversity of malaria parasites circulating in the country is important for establishing a proper control strategy and for understanding the mechanism underlying the pathology of malaria. However, to date no data are available on the genetic diversity of *P. falciparum* in Yemen.

2.7 General glance on previous studies

Many studies have been conducted to investigate the genetic diversity of *P. falciparum* by using the polymorphic regions of merozoite surface proteins *MSP-1* and *MSP-2* as genetic markers in different geographical area in the world as shown in (Tables 2.1 and 2.2). Studies have been done in Myanmar (Kang et al., 2010), Pakistan (Khatoon et al. 2010), Burkina Faso (Soulama et al. 2009), Guinea and Tanzania (Schoepflin et al., 2009), Africa (Bogreau et al., 2006), India (Ranjit et al., 2005), Gabon (Aubouy et al., 2003), Colombia (Gomez et al., 2002; Montoya et al., 2003), French Guiana (Ariey et al., 2001), Uganda (Hoffmann et al., 2001) , Brazil (Silva et al., 2000), Thailand (Snounou et al., 1999), Sengal (Koantae et al., 1999). Extensive parasite diversity is found in highly endemic areas, thus infected human carry multiple genotypes (Babiker et al., 1995; Paul et al., 1998), while the low transmission areas are characterized by limited genetic diversity and most of the infections are monoclonal (Haddad et al., 1999; Hoffmann et al., 2001). In the Eastern Mediterranean Region, studies have been carried out in Iran (Zakeri et al., 2005; Heidari et al., 2007), and Sudan (Babiker, 1995).

Table 2.1 Studies on the genetic diversity of merozoite surface protein (*MSP-1* and *MSP-2*) genes.

Location	Endemicity	*MSP-1*		*MSP-2*		Total	Genetic Diversity	Reference
		MOI	No. alleles	MOI	No. alleles	No. alleles		
Papua New Guinea.	High endemic area	1.99	24	1.84	35	59	Increase genetic diversity in *MSP-2*,	Schoepflin et al. 2009
Tanzania		3.04	76	3.72	29	105	minor increase in *MSP-1*.	
French Guiana	Low endemic region	-	4		2	6	Very limited genetic diversity.	Ariey et al.1999
Burkina Faso	High endemic area	-	-	-	-	70	Differences in Genetic diversity between rural and urban area	Soulama et al. 2009
Iran	Low endemic area	-	9	-	11	20	high polymorphism in a major falciparum malaria endemic region	Heidari et al. 2007
Colombia (Turbo)	Low endemic area	-	1	-	1	2	Population of *P. falciparum* was very homogeneous in both sites	Montoya et al. 2003
Colombia (Zaragoza)		-	3	-	3	6		
Myanmar	High endemic area	2.03	23	2.35	20	43	High degree of genetic polymorphism in *MSP-1* and *MSP-2*	Kang et al .2010
Pakistan	Low endemic region	-	17	-	21	38	High genetic diversity of *P.falciparum.*	Khatoon et al. 2010
Thailand	Low endemic region	1.69	10	2.67	17	27	High genetic diversity of P.*falciparum*	Snounou et al. 1999
Pakistan	Low endemic region	1.25	25	1.22	33	58	Limited genetic diversity	Ghanchi et al. 2010
Honduras	Low endemic area	-	19	-	24	43	Limited genetic diversity	Haddad et al. 1999

Table 2.2 Distribution of the alleles of *MSP-1* and *MSP-2*.

	MSP-1				*MSP-2*			
Location	**K1**	**MAD20**	**RO33**	**Mixed infection**	**FC27**	**3D7/IC**	**Mixed infection**	**Reference**
Papua New Guinea	13	10	1	-	8	27	-	Schoepflin et al. 2009
Tanzania	19	9	1	-	17	59	-	Schoepflin et al. 2009
French Guiana	73	4	20	11	21	74	6	Ariey et al. 1999
Burkina Faso	13	17	11	-	16	13	-	Soulama et al. 2009
Iran	4	4	1	-	6	5	-	Heidari et al 2007
Colombia	-	1	-	-	1	2	-	Montoya et al. 2003
Myanmar	6	17	-	40	8	12	43	Kang et al. 2010
Pakistan	1	14	2	4	6	15	2	Khatoon et al. 2010
Thailand	4	5	1	-	5	12	-	Snounou et al. 1999
Pakistan	12	8	5	-	14	19	-	Ghanchi et al. 2010
Honduras	2	17	-	24	21	-	-	Haddad et al. 1999

CHAPTER III

METHODOLOGY

3.1 Study areas and population

The study was carried out in three malaria endemic regions of Yemen, which were Al-Hudydah, Taizz, and Dhamar, with a total population of 5.9 million. The locations of the study area within Yemen are as shown in Figure 3.1.

The selected regions represent the mountainous hinterland, coastal areas and highland areas, respectively. Living quarters in the rural communities of the study areas are made of mud or stones and have wooden roofs. Majority of the people work in agriculture, fishing, livestock and handicraft sectors.

The peak time of malaria transmission in the coastal areas occurs in winter (October-April), while in the western mountains, the peak occurs during the summer (May-September). The mountainous hinterland normally shows the highest peak of transmission between October and March. In the highlands areas, which are located at more than 2000 metres above sea level, the transmission occurs throughout the year (NMCP, 2002).

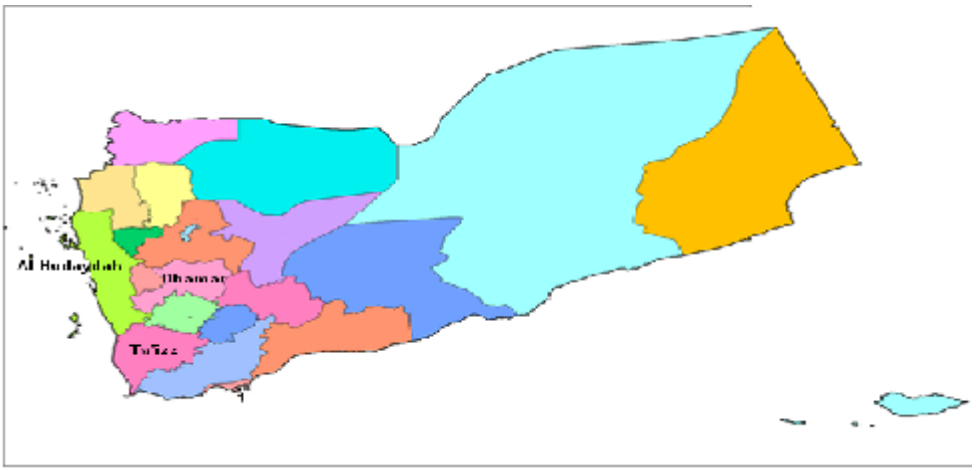

Figure 3.1: Geographical location of study area in Yemen.

Dhamar or Thamar City is a governorate in Yemen. It has a total area of 7,586 km^2 and an estimated population of around 1.5 million (Statistic Year Book of Yemen 2006). It is located to the east of al-Hudaydah. This governorate's climate is temperate, although the central and eastern sections tend to be cold during the winter, while the valleys and western slopes are warmer. The average temperature ranges from 10 to 19°C (50 to 66°F) in the summer, and from 8 to -1 °C (46 to 30°F) in the winter. The governorate lies 1,600–3,200 m (5,200–10,500 ft) above the sea level, with a topographic relief that varies from high mountains to deep valleys, upland plains and plateaus.

Al-Hudaydah is the fourth largest city in Yemen. It is located on the western side of Yemen on the Red sea with a population of 2.3 million people according to the General Census of Population and Housing and Installations for the year 2004. Al-Hudaydah climate is characterized by long hot summer and short cold winter, with average temperature in the summer between (19- 42°C) and (14-24°C) in winter. The rainfall is irregular during the year.

Taizz is a governorate with approximately 11,245 km^2 total area, with long (153 km) seacoast along the Red Sea, with a population of about 2.5 million according to the General Census of Population and Housing and Installations for the year 2004. Taizz is bordered by Al-Hodiedh and Ibb to the north and north east. The weather is fair and throughout the year and it tends to be cold in early morning and late evening.

3.2 Samples and microscopy

Study subjects were patients with febrile illness attending hospitals and medical centres from June 2008 to March 2009. The study was carried out on 511 febrile patients seeking health care. Of these patients, 74 had P. falciparum infection and were selected for the genotyping study. Whole blood from a finger prick was collected from each individual and was used to prepare thick and thin blood films. Additional drops of blood were spotted onto Whatman filter paper 3MM (Whatman International Ltd., Maidstone, England), labelled and stored in plastic bags at room temperature until future use. The thick and thin blood films were stained by Giemsa stain and were examined under microscope at 1000X magnification. Parasitaemia was determined according to the standard techniques and by performing recommended equations (Eliades et al., 2006). Specimens positive for *P. falciparum* were selected for molecular characterization by using *MSP-1* and *MSP-1* markers.

3.3 Genotyping of *P. falciparum* based on *MSP-1* and *MSP-2* genes

3.3.1 Genetic DNA extraction

Extraction of genomic DNA from filter paper blood spot was performed by using Qiagene DNA mini Kit (Qiagene, cat, no. 69581, Germany).

The blood spot on filter paper was cut by flame –sterilized hole puncher and was then put into 1.5 ml microcentrifuge tube, with 20 µl of protien kinase and 180 µl of ATL buffer were then added to the microcentrifuge tube. The microcentrifuge tube was then placed

onto a water bath at 56°C for 1-3 hours for incubation. During the incubation phase, the tube was gently whirled and returned to the heat block every 15 minutes. Then, 200 µl of AL Buffer was added to the tube and was mixed thoroughly by vortexing for 15 sec, then approximately 200 µl of absolute ethanol was later added and the tube was mixed again thoroughly by vortexing for 15 sec. The mixture was pipetted into DNeasy mini spin column in a 2 ml collection tube. The samples were centrifuged at 8000 rpm for 1 min, the column was transferred into another new 2 ml collection tube and then add 500 µl of AW1 buffer, another centrifugation was done at 800 rpm for 1 min and the column was transferred again into a new 2 ml collection tube, 500 µl AW2 buffer was added and centrifuged at 14000 rpm for 3 min. The column was finally transferred into 1.5 ml microcentrifuge tube and the final addition was done by adding 50-200 µl of AE elution buffer and then centrifuged at 8000 rpm for 1 min. The final filtered product was conserved in a new labeled tube and stored at -20°C until it was used for the amplification reaction.

3.3.2 Nested PCR amplification

Nested PCR was carried out according to Snounou et al., (1999). In the first amplification reaction, oligonuclueotide primers that annealed to the conserved region flanking the repeat polymorphic sequences of the genes were used. In the Nest-2 PCR in which another pair of oligonuclueotide primers were used to recognize the sequences that contained the specific allele. Polymorphic block 2 of *MSP-1* can occur as one of three distinct families i.e., K1, MAD20, or RO33. Different repeated sequence units characterize the K1 and MAD20 families, whereas the RO33 block 2 regions consist of a unique sequence. The sequences in block 2 that flank the repeated region of K1 or

MAD20 are unique to each family and are shared amongst all the allelic variants. Three separate Nest 2 reactions were therefore performed to complete the genotyping using the *MSP-1* marker (Table 3.1): specific for the K1 family [(M1–KF) + (M1–KR)], MAD20 family [(M1–MF) + (M1–MR)], and RO33 family [(M1–RF) + (M1–RR)].

Polymorphic block 3 of *MSP-2* can occur as one of the two distinct families i.e., FC27 and 3D7/IC. Different repeated sequence units that characterized the FC27 and 3D7/IC families and the sequences in block 3 that flank these repeated regions are unique to each family and are shared amongst all the allelic variants. Two separate Nest 2 reactions are therefore performed to complete the genotyping of parasites using the *MSP-2* marker (Table 3.1): One specific for the FC27 family [(M2–FCF) + (M2– FCR)] and one specific for the 3D7/IC family [(M2–ICF) + (M2–ICR)].

3.1 The primer sequence for both *MSP-1* and *MSP-2* alleles.

MSP-1	*Sequence*
M1-OF	5- -CTAGAAGCTTTAGAAGATGCAGTATTG-3-
M1-OR	5- -CTTAAATAGTATTCTAATTCAAGTGGATCA-3-
M1-KF	5- - AAATGAAGAAGAAATTACTACAAAAGGTGC-3-
M1-KR	5- -GCTTGCATCAGCTGGAGGGCTTGCACCAGA-3-
M1-MF	5- -AAATGAAGGAACAAGTGGAACAGCTGTTAC-3-
M1-MR	5- -ATCTGAAGGATTTGTACGTCTTGAATTACC-3-
M1-RF	5- -TAAAGGATGGAGCAAATACTCAAGTTGTTG-3-
M1-RR	5- -CATCTGAAGGATTTGCAGCACCTGGAGATC-3-
MSP-2	
M2-OF	5- -ATGAAGGTAATTAAAACATTGTCTATTATA-3-
M2-OR	5- -CTTTGTTACCATCGGTACATTCTT-3-
M2-FCF	5- -AATACTAAGAGTGTAGGTGCARATGCTCCA-3-
M2-FCR	5- -TTTTATTTGGTGCATTGCCAGAACTTGAAC-3-
M2-ICF	5- -AGAAGTATGGCAGAAAGTAAKCCTYCTACT-3-
M2-ICR	5- -GATTGTAATTCGGGGGATTCAGTTTGTTCG-3-

For all the PCR amplifications, the total volume for each reaction was 50 µl. A master mix containing all the reagents, except for the DNA, was prepared and aliquoted into the reaction tubes. All reagents were fully thawed and vortexed before being used to prepare the master mix. The DNA template was added last. Primary PCR was performed in 50 µL reaction mixture containing 5 µL of DNA template, 1X i-Taq^{TM} buffer free of $MgCl_2$ (iNtRON BIOTECHNOLOGY, Seoul, Korea), 1mM of $MgCl_2$ (iNtRON BIOTECHNOLOGY, Seoul, Korea), 125 µM dNTP (iNtRON BIOTECHNOLOGY, Seoul, Korea), 0.25 µM each of the outer primer pairs *MSP-1* Forward(F)/ Reverse (R) and *MSP-2* (R/F) and 1.0 U of i-Taq^{TM} DNA polymerase (iNtRON BIOTECHNOLOGY, Seoul, Korea). Secondary PCR was performed in 50 µL reaction mixtures containing 3 µL of DNA template with the same concentrations as the primary PCR. In both amplifications, samples were incubated in the thermo cycler (Bio-Rad, Hercules, USA) under the following conditions: initial denaturing step at 95ºC for 5 min, annealing at 58ºC for 2 min and extension at 72ºC for 2 min, followed by 24 cycles of denaturing for 1 min at 94ºC, annealing for 2 min at 58ºC and extension for 2 min at 72ºC, followed by a final annealing at 58ºC for 2 min and a final extension at 72ºC for 5 min. The PCR products were subjected to electrophoresis in 1.5% agarose gels and stained with Sybr green. The sizes of *MSP-1* and *MSP-2* alleles were measured using Quantity One software (Bio-Rad, USA). If the sizes were different by approximately up to 10 bp, they were still considered as the same and categorized as one allele.

3.4 PCR product analysis

3.4.1 Gel preparation

Each PCR product was electrophoresed on a 1.5 % agarose gel which was prepared by adding 0.6 g of agarose to 100 ml flask containing 40 ml of 1X TAE buffer. The agarose was melted thoroughly by heating in a microwave oven at high power for 1-2 min. Then, the dissolved agarose mixture was allowed to cool to approximately $50C^0$. Subsequently, 0.5 ug/ml final concentration of Sybr safe (Invitrogen) was added and the gel poured into a mould, fitted with appropriate comb for casting. The gel was allowed to set for at least 30 min to solidify and was later was placed in tank containing 1X TAE buffer.

.

3.4.2 Gel electrophoresis

An aliquot of 2 µl of 6X gel loading buffer (Fermentas) was mixed with 10µl of each amplifed product on a sheet of clean parafilm. 6 µl of the mixture was transfered to the respective well. DNA molecules standard (100bp) was also loaded into a well in the gel.

Electrophoresis was run at 110 V for 40 minutes and visualization of Sybr safe stained DNA was performed by using gel documentation system. The sizes of *MSP-1* and *MSP-2* alleles were measured using Quantity One software (Bio-Rad, USA).

The detection of a single PCR fragment of an allelic family was considered as an infection with one genotype. Multiple infection of *P. falciparum* was defined as the presence of more than one genotype of either *MSP-1* or MSP-2 in a single blood sample

(Soulama et al. 2009). The complexity of *P. falciparum* (multiplicity of infection) is the mean of the frequency of the multiple infections in a single blood sample i.e., the total of genotypes divided by the total number of samples (Soulama et al. 2009). Complexity and multiple infections were calculated by combining the *MSP-1* and MSP-2 results.

3.5 Statistical analysis

Data analysis was performed using the SPSS programme for Windows version 11.5 (SPSS Inc., Chicago, IL, USA), and associations between the proportions were tested using the χ^2 test. T-test and One Way ANOVA were used to compare between the means and significance was defined as $P < 0.05$

3.6 Ethical clearance

Patients participated on a voluntary basis after they were given a clear explanation of the research objectives. If the subjects were children, informed consent was obtained from their guardians. Study protocol was reviewed and approved by the Faculty of Medicine, Sana'a University, Sana'a, Yemen.

CHAPTER IV

RESULT

4.1 Characterization of study population and the distribution of P. faclciparum

This study was carried out on 74 patients who had *P. falciparum* infection. Of which, 60.8% were males and 39.2% females. The distribution of gender in the three different locations is indicated in Figure 4.1. The prevalence (Figure 4.2) of *P. falciparum* in the three geographical regions were 17% (Taizz), 49% (Dhamar) and 34% (Al-Hudaydah).

Parasite density (Figure 4.3) varied considerably among the study areas, ranging from 300-150,000 parasites/µl (mean 32422.68). The density of parasite in patients from Taizz ranged between 22000 and 64,000 parasites/µl (mean 35,627.27), while in Dhamar it was between 700 and 98,000 parasites/µl (mean 37,772.95), and in Al-Hudaydah it ranged from 300 to 150,000 parasites/µl (mean 22,438.89).

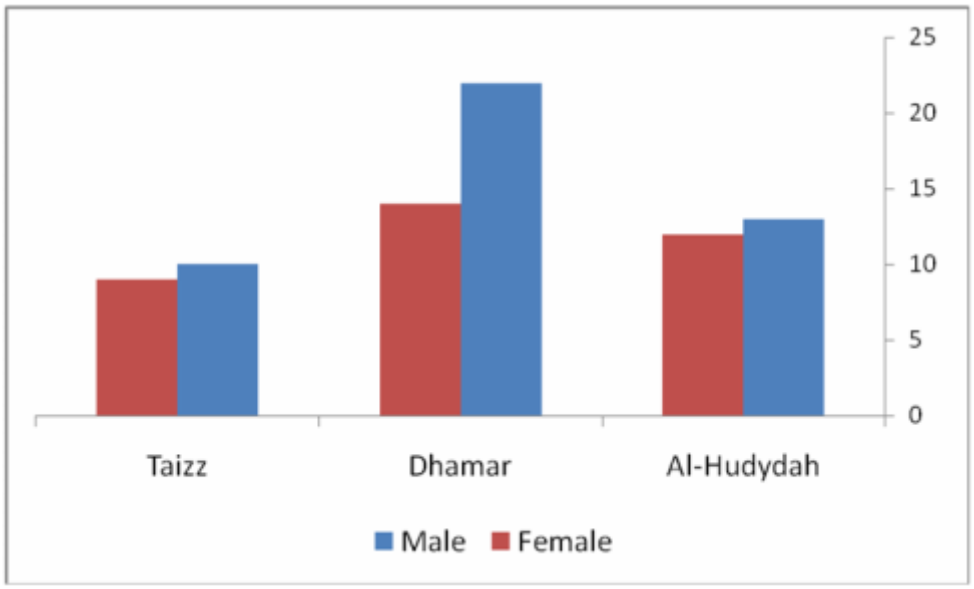

Figure 4.1: The distribution of participants according to gender in the three locations

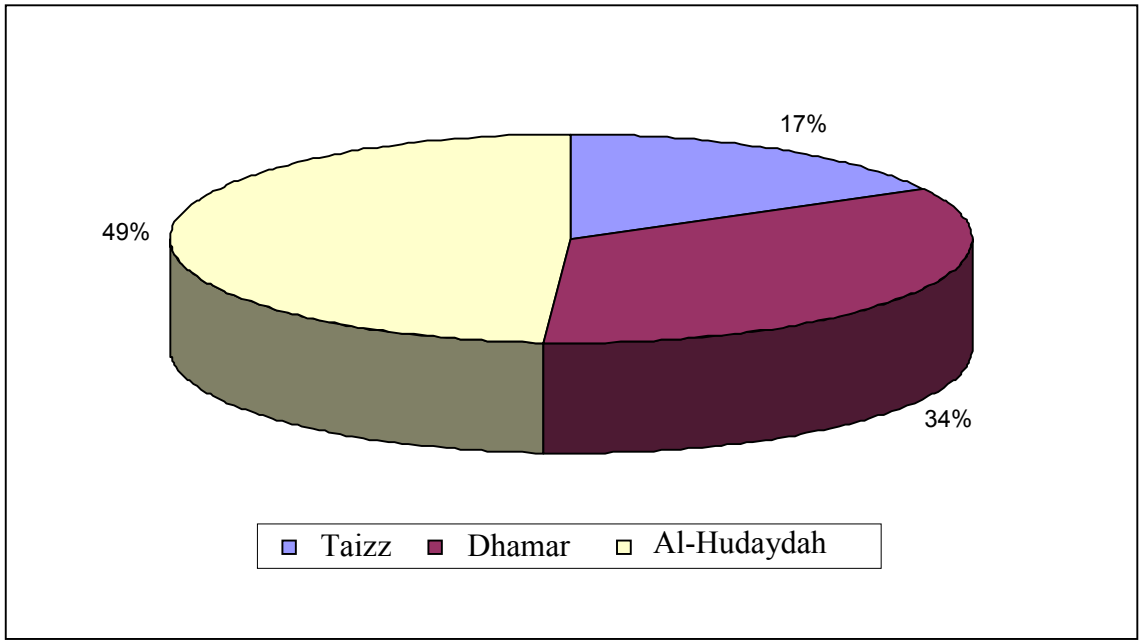

Figure 4.2: Percentage of samples collected from study areas (Taizz, Damar, and Al-Hudaydah) in Yemen.

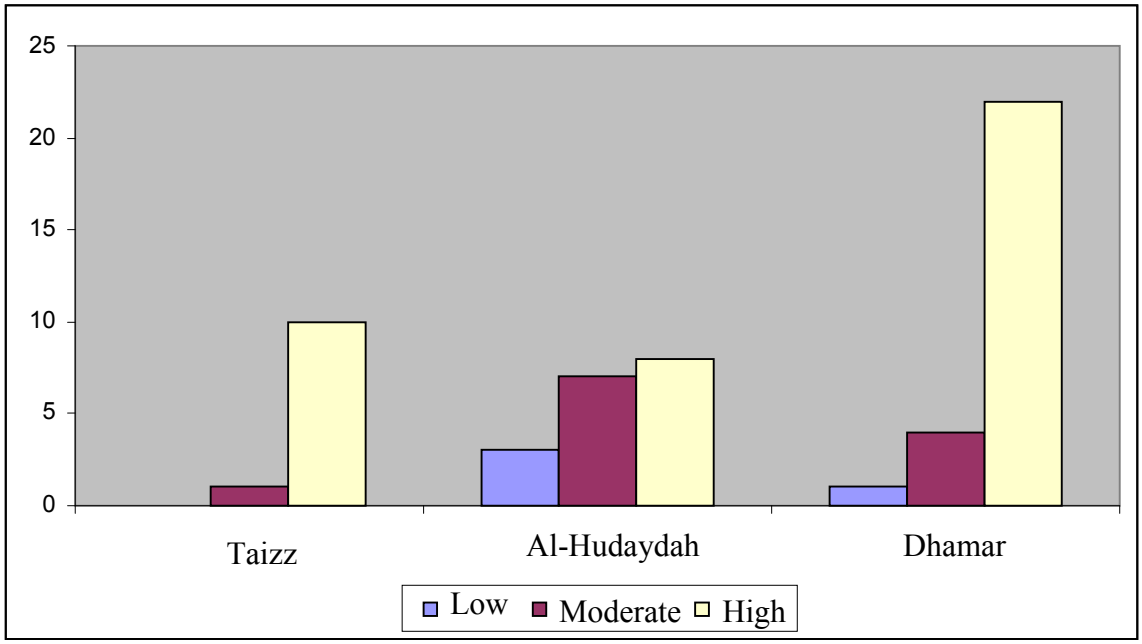

Figure 4.3: Parasite density among patients from the three study locations.

4.2 Nested PCR

Nested PCR produced amplicons of different sizes based on the allelic types from all samples. Primers specific for K1 allelic family of *MSP-1* produced amplicons with sizes ranging from 100 bp to 230 bp, representing different types of alleles (Figures 4.4 and 4.7). The sizes of MAD20 amplicons ranged from 130 bp to 230 bp (Figures 4.5 and 4.8). The range of amplicon sizes of RO33 was 130 bp to 190 bp (Figures 4.5 and 4.9). High range of sizes of amplicons was detected among allelic families of *MSP-2* (Figures 4.6, 4.11, and 4.12).

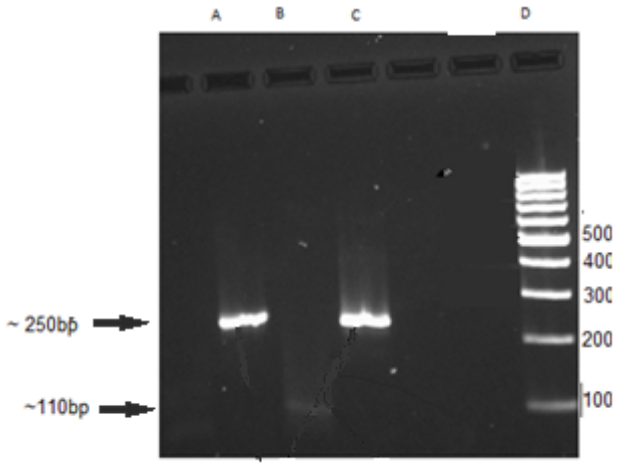

Figure 4.4: Different allelic forms (100 and 250 bp alleles) of K1. Lane D represents the molecular weight marker (100bp) and lanes A–C are *P. falciparum* samples from patients.

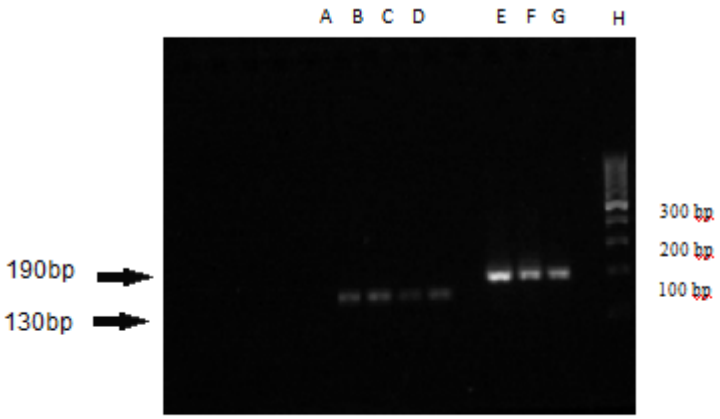

Figure 4.5: Different allelic forms of RO33 and MAD20 detected (130, and 190 bp alleles respectively). Lane H represents the molecular weight marker (100bp), lanes A–C represent MAD20, and lanes E-G are RO33 of *P. falciparum* samples from patients.

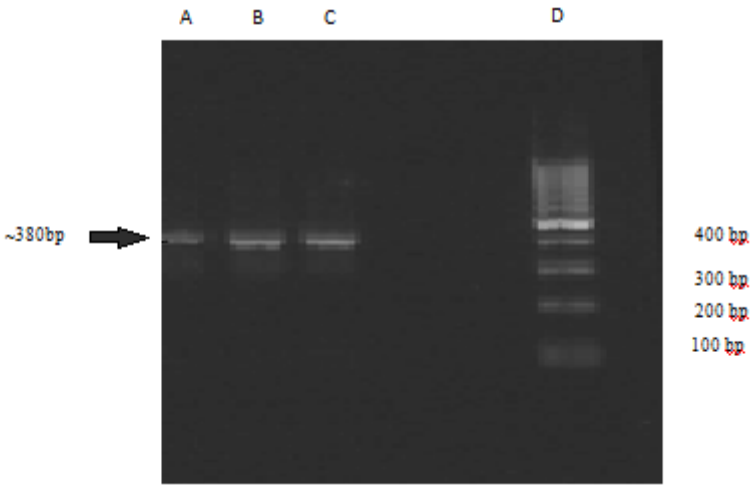

Figure 4.6: FC27 allelic forms of *MSP-2* (400 bp). Lane D represents the molecular weight marker (100bp) and lanes A–C are *P. falciparum* samples from patients.

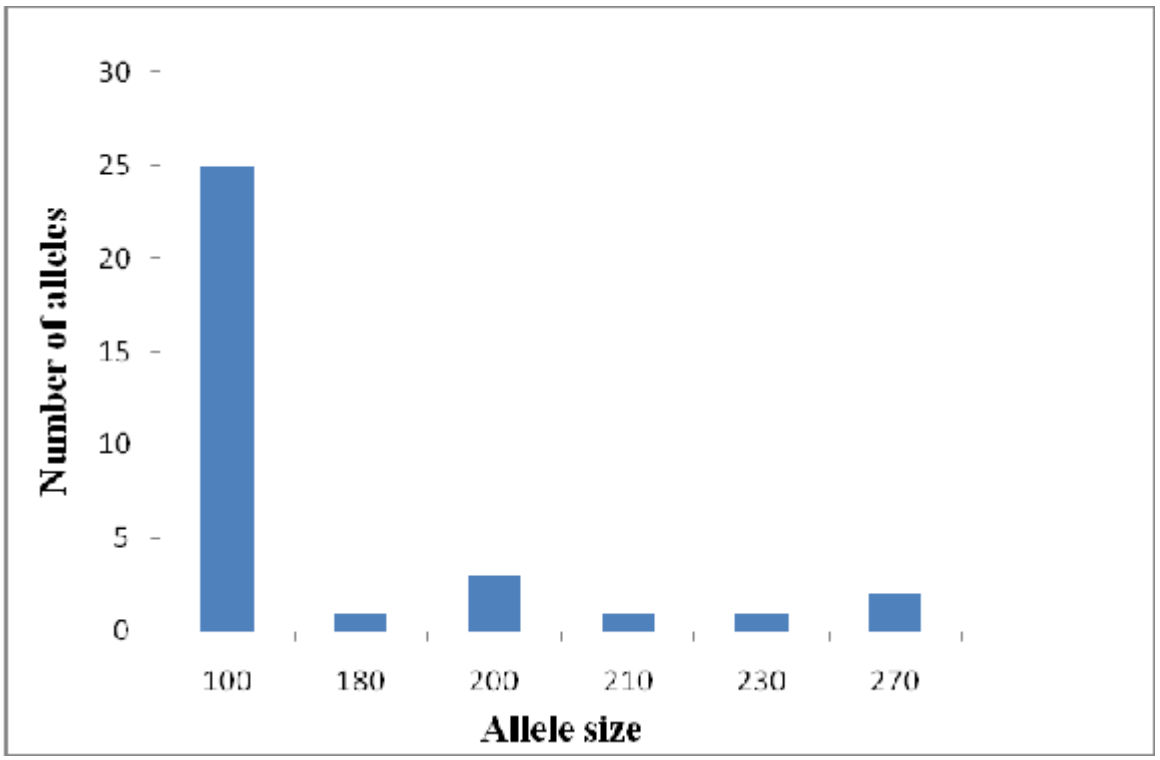

Figure 4.7: Size and frequency of K1 allele detected in the study areas.

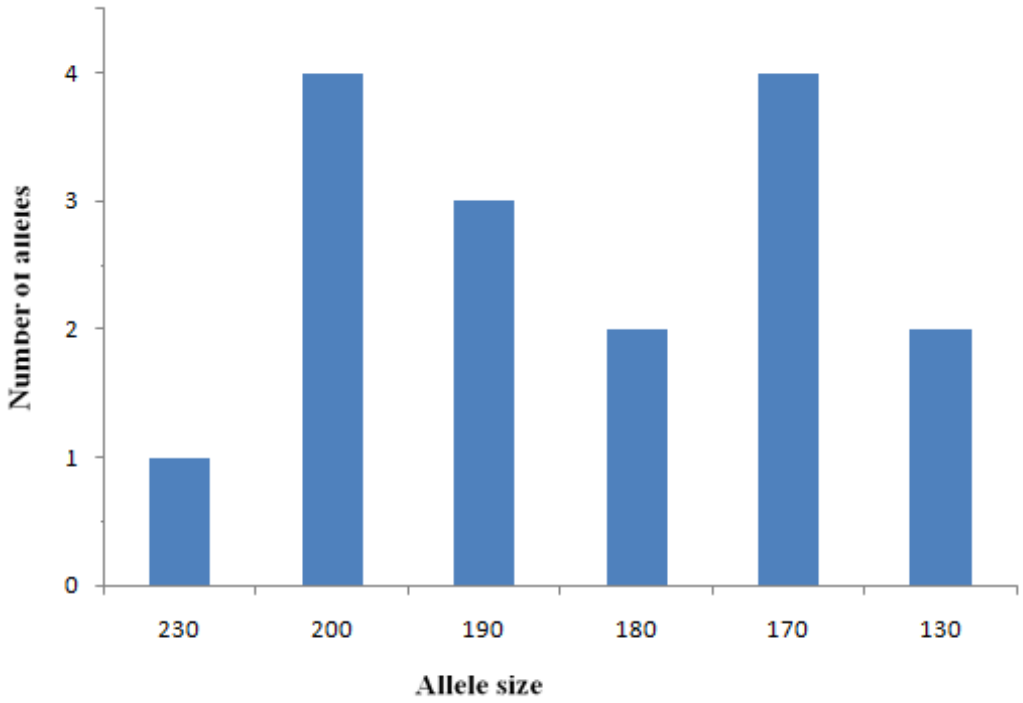

Figure 4.8: Size and frequency of MAD20 observed in the study areas.

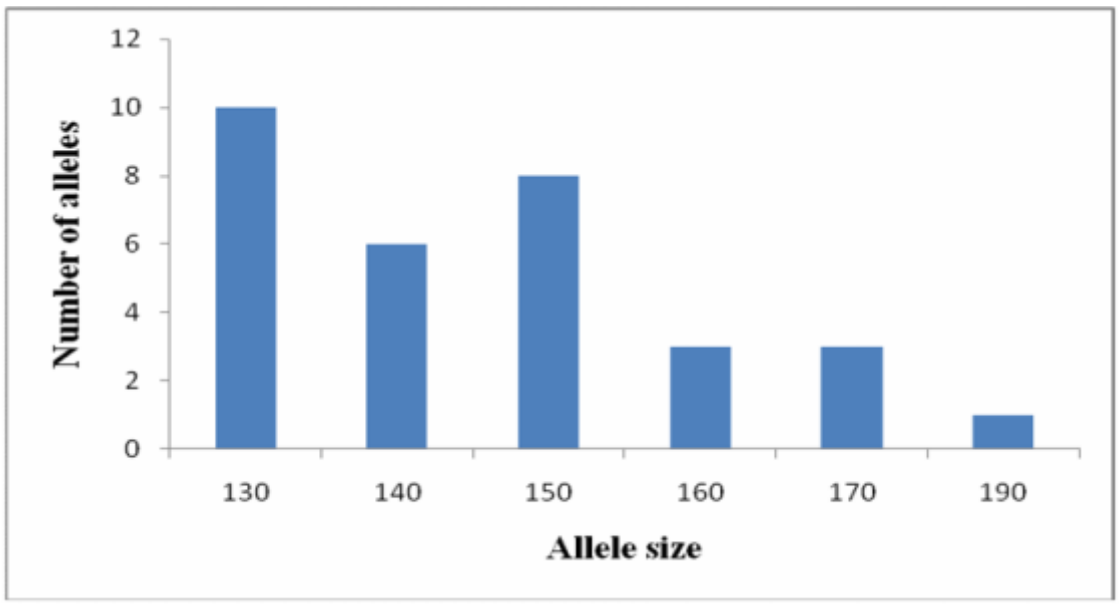

Figure 4.9: Size and frequency of RO33 detected in the study areas.

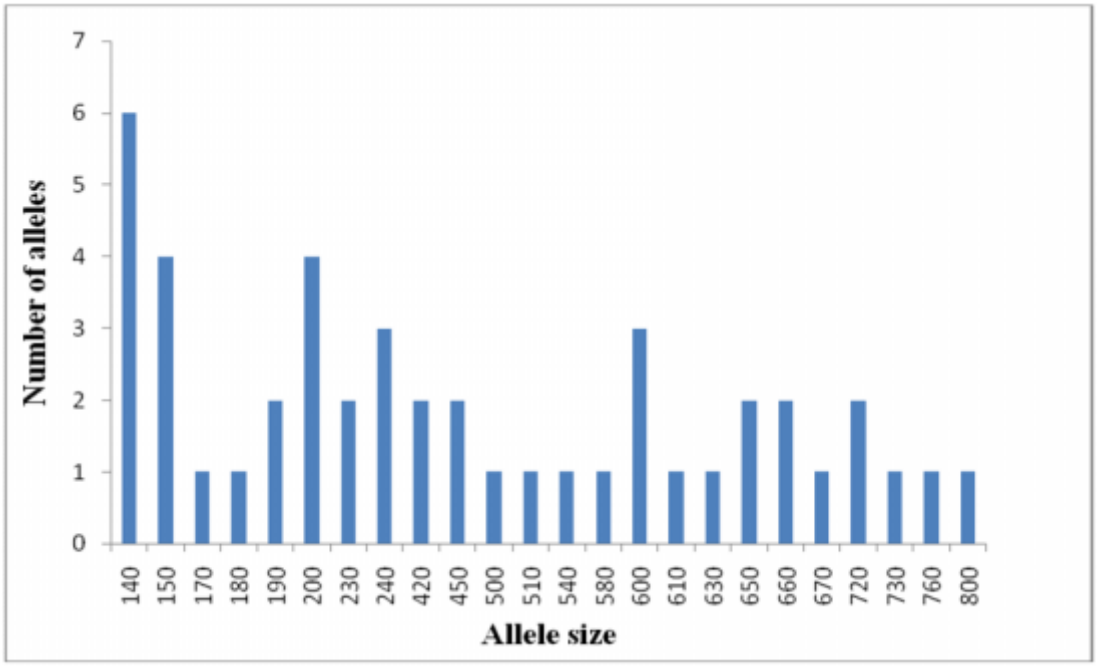

Figure 4.10: Size and frequency of 3D7 detected in the study areas.

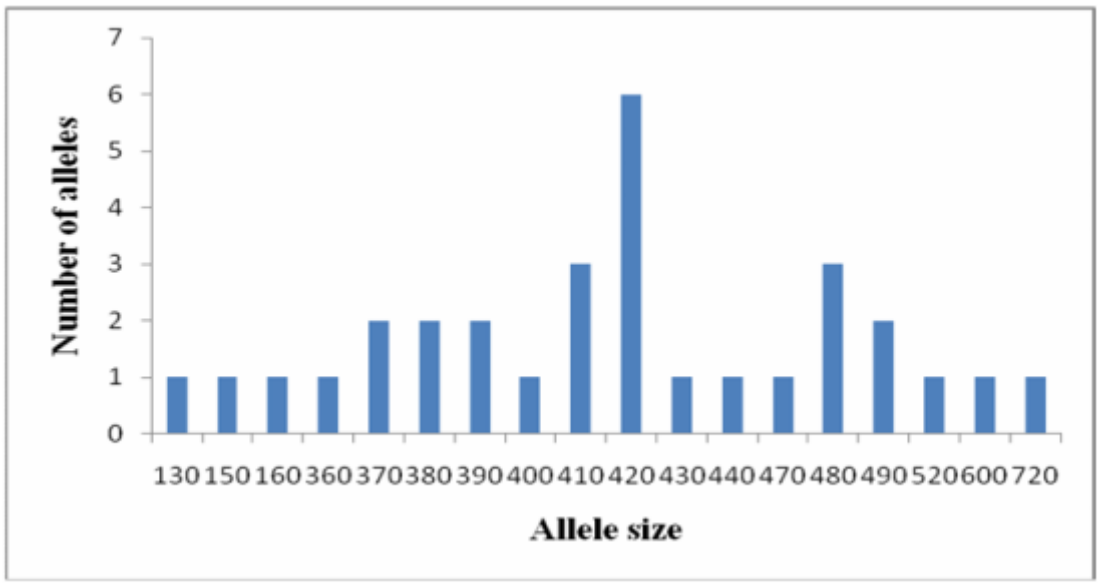

Figure 4.11: Size and frequency of FC27 observed in the study areas.

4.3 Prevalence and distribution of allelic families of *MSP-1* and *MSP-2*

All samples were analyzed for polymorphism on *MSP-1* and *MSP-2*. The various allelic forms of *MSP-1* and *MSP-2* were classified according to their size. Among the fragments of 74 samples at the *MSP-1* locus, 33, 31 and 16 fragments belonged to K1, RO33 and MAD20 families, respectively. For the *MSP-2* locus, 48 and 31 fragments were identified as 3D7 and FC27 families, respectively (Table 4.1). Seventy six individuals of MSP alleles were identified in all the study areas; 7 different alleles for K (100-270 bp); 10 for MAD20 (130-230 bp); 11 for RO33 (130-190 bp); 22 alleles for FC27 (130-720 bp) and 26 alleles for 3D7. However, the variant 100 bp of K1, 200 bp and 170 bp of MAD20 alleles, and 140 bp of RO33 family *MSP-1*, and the variant 420 bp of FC27 and 140 bp of IC for *MSP-2* demonstrated the highest frequency (Figures 4.7-4.11). For *MSP-1*, the RO33 allele family was the most diversed with 11 different alleles, followed by MAD20 with 10 alleles diversed, the K allele was the least diversed with 7 alleles. For the *MSP-2*, the 3D7 allele family was the most diversed with 24 different alleles, compared to 18 from the FC27 allele family.

Of all the three study areas, the total number of *MSP-2* detected was greater than *MSP-1* with 48 and 28 alleles, respectively. As shown in Table 4.2, 3D7 (68 %) allelic family was the most predominant, followed by K1 (45%), RO33 (42%), FC27 (42%) and MAD20 (22%). For the *MSP-1* alleles, the most predominant in Dhamar was RO33 followed by MAD20 and only one K alleles. While in Al-Hudaydah, K1 and MAD20 were present at the same frequency. The RO33 and MAD20 were present at the same frequency in Taizz.

The frequency of mixed infection with more than one family of alleles were 32% for FC27+ 3D7 and 18% for MAD20 + RO33. The four allelic families (MAD20, RO33, FC27 and 3D7) were significantly more prevalent in Taizz areas as compared to Al-Hudaydah and Dhamar. In contrast, the K1 allele was most frequent in Dhamar. A total of 55 patients (77.5%) had more than one allele. There was no significant differences in prevalence of allelic families between rural and urban areas.

Table 4.1**:** Distribution of families of *MSP-1* in Yemen *P. falciparum* isolates

Marker	Prevalence n (%)					
	Al-Hudaydah	Dhamar	Taizz	Total	χ^2	*P* value
MSP-1	n= 25	n= 36	n= 13	n= 74		
K1	8 (32)	22 (61)	3 (23)	33 (45)	8.016	< 0.05
MAD20	8 (32)	2 (6)	6 (46)	16 (22)	11.689	< 0.01
RO33	15 (60)	6 (16.7)	10 (77)	31 (42)	19.332	< 0.001
K1 + MAD20	3 (12)	1 (3)	0 (0)	4 (5)	-	-
K1 + RO33	5 (20)	2 (6)	1 (8)	8 (11)	-	-
MAD20 + RO33	6 (24)	1 (2.8)	6 (46)	13 (18)	13.408	< 0.001
K1 + MAD20 + RO33	1 (4)	0 (0)	0 (0)	1 (1.4)	-	-

Table 4. 2: Distribution of families of *MSP-2* in Yemen *P. falciparum* isolates

Marker	Prevalence n (%)					
	Al-Hudaydah	Dhamar	Taizz	Total	χ^2	*P* value
MSP-2	n= 25	n= 36	n= 13	n= 74		
FC27	12 (48)	9 (25)	10 (77)	31 (42)	11.157	*< 0.01*
3D7/IC	16 (64)	21 (58)	13 (100)	50 (68)	7.786	*< 0.05*
FC27 + 3D7/IC	10 (40)	4 (11)	10 (77)	24 (32)	19.864	*< 0.001*
Mixed infection*	18 (72)	24 (72.7)	13 (100)	55 (77.5)	4.634	*> 0.05*

4.4 Complexity of families of *MSP-1* and *MSP-2* in Yemen *P. falciparum* isolates

The complexity of multiple infections (MOI) was defined as the mean of the frequency of multiple infections. The overall complexity was 2.18. The complexity of infection showed significant difference between areas (P = 0.001, F = 14.56) where the highest complexity was observed in Dhamar followed by Al-Hudaydah and Taizz, Urban areas had higher complexity of infection as compared to rural areas (t = 2.18, $p < 0.05$). ANOVA test showed no significant association between MOI with age, gender and parasitaemia (Table 4.3).

Table 4.3 complexity of *P. falciparum* based on *MSP-1* and *MSP-2* in Yemen isolate.

Variables	Mean ± SD	95%CI	Significance
Study area*			
Hinterland	3.23 ± 0.83	2.72 - 3.73	F=14.56, P <0.001
Coastland	2.36 ± 1.15	1.88 - 2.83	
Highland	1.67 ± 0.75	1.41 - 1.92	
Location#			
Rural	2.08 ± 1.06	1.78 - 2.39	T = 2.18, P < 0.05
Urban	2.90 ± 1.10	2.11 - 3.69	
Age*			
< 5	2.5 ± 0.86	1.74 - 2.56	F = 1.014, p > 0.05
5-10	1.9 ± 1.46	2.07 - 2.92	
> 10	2.15 ± 1.14	0.65 - 3.09	
Parasitaemia*			
Low	1.75 ± 0.95	0.23 - 3.27	F = 1.066, p > 0.05
Moderate	2.00 ± 0.73	1.53 - 2.47	
High	2.35 ± 1.21	1.96 - 2.74	

* Difference tested using one Way ANOVA,

Difference tested using t-test.

CHAPTER V

Discussion and Conclusion

5.1 Discussion

This study was carried out to determine genotypes and population structure of *P. falciparum* in Yemen using the two most polymorphic regions of *MSP-1* and *MSP-2* genes. The genetic diversity of *P. falciparum* was compared in three different geographical areas endemic for malaria. No such study had been undertaken to date in Yemen, and the need for detailed studies of the nature and extent of *P. falciparum* diversity in this area is therefore obvious.

Genetic structure of *P. falciparum* populations has been shown to play a role in the natural acquisition of immunity in malaria infection (Healer et al., 2004), Thus, knowledge of the genetic structure of *P. falciparum* population is necessary to develop strategies for controlling the disease, including the design of effective vaccine against *P. falciparum,* and also to prevent the rapid spread of parasites resistant to drugs or vaccine (Tibayrenc, 1998).

Results show high proportion of all the allelic families (K1, MAD20, RO33, FC27, and 3D7), with 3D7 being the most frequent and MAD20 the least. Within the families of *MSP-1* gene, the K1 (33/74, 45%) was the more frequent allelic family, followed by RO33 (31/74, 42%), while MAD20 (16/74, 22%) was less frequent. These findings are

similar to studies carried out in French Guinea (Ariey et al., 1999), Kenya (Takala et al., 2002), and Peru (Chenet et al., 2008). On the other hand, the findings are not consistent with studies that were carried out in Thailand (Snounou et al., 1999), Pakistan (Ghanchi et al., 2010), Iran (Zakeri et al., 2005), Myanmar (Kang et al., 2010), and Colombia (Gomez et al., 2002). In these studies, it was found that MAD20 was the predominant allele.

Analysis of *MSP-2* (block 3) showed higher prevalence of 3D27/IC allelic family than FC27 with high proportion of mixed infection with the two families. Similar frequency was found in Iran (Zakeri et al., 2005), Cameroon (Basco et al., 2004) Pakistan (Ghanchi et al., 2010), Myanmar (Kang et al., 2010), and Thailand (Snounou et al., 1999). However, the findings in this study are not similar to those studies carried out in Brazil (Sallenave et al., 2003), and Burkina Faso (Soulama et al., 2009), in which FC27 was more prevalent. The difference in genetic diversity of *P. falciparum* populations could be explained by the high malaria transmission intensity in the different malaria areas (Babiker et al., 1995, Paul et al., 1998). A high endemic area is generally characterized by extensive parasite diversity and infected humans often carry multiple genotypes. Conversely, the parasite population in a low transmission area has limited genetic diversity and the majority of infections are monoclonal (Haddad et al., 1999; Babiker et al., 1997; Gomez et al., 2002; Hoffmann et al., 2001). The four allelic families (MAD20, RO33, FC27 and 3D7) were significantly more prevalent in the hinterland areas as compared to coastland and highland areas. In contrast, the K1 allele was most frequent in the highlands. Although the allelic families of *MSP-1* and *MSP-2* were not restricted to

geographical location, a significant difference was noted between areas and this could be attributed to the different levels of endemicity in the three locations.

The present study has shown that *MSP-2* locus was more diverse, with a total of 42 alleles (68%) as compared to18 (32%) *MSP-1* alleles. This observation is in agreement with results reported in Senegal (Konate et al., 1999), Tanzania (Smith et al., 1999), Papua New Guinea (Felger et al., 1994), Sudan (Babiker et al., 1997), Colombia (Snewin et al., 1991), and Honduras (Haddad et al., 1999). The *MSP-2* gene has been considered the most informative and robust marker for assessing genetic diversity of *P. falciparum* (Farnert et al., 2001). The positive correlation between numbers of *MSP-2* alleles and endemicity of malaria has been widely reported (Babiker et al., 1997; Felger et al., 1994).

Mixed infection was observed in 77.5% of patients and almost all patients from the hinterland areas harbored more than one genotype. Mixed infection rates of almost 50% have been observed in mesoendemic regions (Conway et al., 1991; Zwetyenga et al., 1998), and up to 100% in holoendemic regions (Babiker et al., 1997; Smith, et al., 1999). In areas with low endemicity, mixed infections can range from 20-30% (Babiker et al., 1997; Haddad et al., 1999). Meiotic recombination during the sexual reproduction stage is a major mechanism for the generation of genetic diversity of *P. falciparum*. Thus, the high prevalence of mixed infection can enhance cross-fertilization resulting in the generation of novel genotype combinations and novel alleles (Vafa et al., 2008).

The complexity of *P. falciparum* infection in Yemen was 2.18 which is almost similar with the complexity reported previously from hyperendemic areas (Babiker & Walliker 1997). The highest complexity was observed in the hinterland areas, followed by the coastland and highland areas. These differences are statistically significant, suggesting that endemicity and level of malaria transmission were different in the three locations. The positive association between complexity and endemicity of *P. falciparum* has been well documented (Babiker et al. 1997; Babiker & Walliker 1997; Paul & Day 1998; Paul et al. 1998) with few exceptions in certain areas where high complexity was reported from low endemic areas (Sakihama et al. 2006; Zakeri et al. 2005). The complexity in the hinterland and the coastland was similar to those observed in hyperendemic areas in Africa (Babiker et al., 1997; Babiker & Walliker, 1997; Schoepflin et al., 2009), while the complexity in the highland was similar to those reported from low transmission areas such as in central and southeast Asian countries (Ghanchi et al., 2010; Joshi et al., 2007; Paul et al., 1998).

It was noted that complexity of falciparum malaria was unexpectedly higher in urban as compared to rural areas. However, this comparison may be biased by the fact that complexity in the urban areas was based on a smaller sample size as compared to that of the rural areas (10 *vs* 40). Nonetheless, the high complexity in the urban areas may be explained by the massive migration of people from the rural of Yemen to the city. An individual may acquire the infection in the rural area before moving to the urban area, and this infection may be overlapped by a new infection. It has been reported that *P. falciparum* infection can persist for more than 2 years in a semi-immune patient (Krajden

et al., 1991). The increase in the complexity due to the migration between areas of more intense transmission was postulated in a study from Southeast Asia (Paul et al., 1998).

In the present study, no positive association was observed between complexity and age. Similar findings have also been seen in previous studies (Felger et al., 1999; Issifou et al., 2001; Paul et al., 1998; Zwetyenga et al., 1998). However, a positive association between complexity and age was reported from the rural areas of Burkina Faso (Soulama et al., 2009; Soulama et al., 2006). Significant association between complexity and parasite density was documented from southern Mozambique (Mayor et al. 2003); Senegal (Vafa et al., 2008) and Tanzania among infants but not in adults (Felger et al., 1999). In this study, however, there was no significant association between complexity and parasite density.

5.2 Conclusion

This study found high genetic diversity among *P. falciparum* populations in Yemen. Most of the infections were mixed with a high level of MOI. The significant difference in complexity and the distribution of the family alleles between hinterland, coastland and highland areas suggests the difference in malaria transmission between these areas. This high genetic diversity of *P. falciparum* Yemen represents a major challenge for developing malaria control strategies. Furthermore, high genetic diversity may provide a mechanism for the emergence of drug resistance in the *P. falciparum* populations.

5.3 Recommendation

This study is the first attempt to establish molecular techniques for genetic diversity of *Plasmodium falciparum* in Yemen. Further studies is necessary to include all endemic provinces of Yemen, moreover future studies should also focus on the use of different markers for diversity such as microsatellites.