Jos. J. Schall
University of Vermont
30 December 2005

Isolation of microsatellite markers from the genome of malaria parasites

The malaria parasites are an extremely diverse and geographically wide-spread group. These parasites have been placed into two genera, *Plasmodium* and *Haemoproteus*. Although the six species infecting humans (four well-known, plus one recently discovered, and one thought to be a monkey parasite is actually common in humans in some locations) are best known to most biologists, there is a very large number of species infecting other vertebrate hosts. Most of these are parasites of birds and reptiles.

Population genetic studies, and studies on the clonal diversity of infections, require variable genetic markers for the parasites. Full-genome sequencing of the important malaria parasite of humans (*P. falciparum*) led to the discovery of thousands of microsatellite markers, and these are now being used in important studies in the basic biology of the parasite. Unfortunately, *P. falciparum* is only very distantly related to the malaria parasites of birds and reptiles, so the described primers do not amplify homologous targets for the parasites of nonhumans. (We tried 20 pairs with the lizard malaria parasite, *P. mexicanum*, without luck, although all pairs worked very well for a sample of *P. falciparum*.)

Discovery of microsatellites is fairly routine, using an enrichment procedure that is now standard in many labs. However, any sample of vertebrate blood infected with *Plasmodium* will contain a mix of host and parasite DNA. For the malaria parasites of lizards and birds, this problem is particulary acute because these vertebrates have nucleated erythrocytes (mammal red blood cells lack a nucleus). Thus, an infected reptile or bird will have far more host DNA in its blood than parasite DNA. The *Plasmodium* genome is about 24 million bases in size, much smaller than that of a vertebrate. The parasite in the vertebrate is haploid, thus will have half the DNA of a similar diploid cell. Also, most infections do not include parasites in a high proportion of host red blood cells. In a bird or lizard host, there will be hundreds of times more host than parasite DNA. This simple fact has hindered discovery of microsatellite markers for the avian and saurian malaria parasites.

We have perfected a protocol that solves this problem. The methods are described in detail below, and we include sources for all of our reagents in **bold** (there are certainly many other possible sources). The protocol can be used for a standard search for microsatellites for any species (not just for

malaria parasites), and we note changes that would be recommended at the end of this document.

The method depends on an important feature of the genome of some (most?) *Plasmodium* species, their high At-richness. Full-genome sequencing reveals that *P. falciparum* [human parasite] and *P. yoelii* [rodent parasite] have about 80% AT in their genomes (the proportion varies within the genome). A similar high AT-richness is known for another rodent parasite, *P. berghei*. Partial sequencing has been done for the genome of the bird malaria parasite, *P. gallinaceum*, and the results can be inspected on the Sanger site. Again, the AT percentage of bases is about 80%. The sequence of *P. reichenowi*, a parasite of chimps, is also at 80% AT. The one *Plasmodium* that does not fit this picture (at least what is known), is *P. vivax*, a common human malaria parasite, which has a "normal" AT percent of about half of the bases. We proceeded with our study of a lizard malaria parasite with the hope that this parasite also has a high AT-richness, and this seems to be the case.

These methods were used to find microsatellites in the genome of the lizard malaria parasite, *Plasmodium mexicanum*, a parasite of the western fence lizard, *Sceloporus occidentalis*. They should work for other species of *Plasmodium*, and likely also for *Haemoproteus*. Again, with the hope that these parasites are more like *P. falciparum*, and not like *P. vivax*. The methods are worth a try for the other "bird malaria" genus, *Leucocytozoon*, but the AT-richness of that parasite's genome is not known (only the mitochondrial cytochrome b gene and the SSU rRNA gene have been sequenced for that genus).

STEP 1. Extracting DNA from host blood

Fresh blood from heavily infected lizards was extracted using a GenElute Miniprep DNA extraction kit for mammalian blood (**Sigma, St. Louis**), and the protocol supplied with the kit.

High parasitemia in the blood (percent of red blood cells infected) will yield a higher concentration of parasite DNA in the extract. This is obvious. But, how high must this parasitemia be for success? We used infections that ranged from about 50 - 80% parasitemia, which is very high even for *P. mexicanum*. Although not tested, the methods described here will most likely work even with lower concentration of parasite DNA.

For the digest below, there should be about 30,000 ng of DNA in the extract. We used a Nanodrop spec to determine the concentration. If the total amount of DNA is less than this, the digest volume can be reduced appropriately.

If the extracted DNA is concentrated, use a vacuum centrifuge with the heat OFF to prevent any damage to the DNA. Watch closely to avoid drying the DNA all the way down, which would make it difficulty to get it back into solution.

STEP 2. Restriction digest of the DNA

There are hundreds of restriction enzymes available. For *Plasmodium* DNA, the enzyme should be one that cuts at CT regions of the genome. Any enzyme that cuts around AT areas will dice the parasite's genome into very small pieces! (remember that *Plasmodium* genome is AT rich!) Also, a six base cutter will result in fewer small fragments.

We used the Sau-AI enzyme, a four base cutter, with its buffer kit from **New England BioLabs (NEB, Beverly, MA)**. This enzyme cuts:

NNNNNNNNN>GATCNNNNNNNNN
NNNNNNNNNCTAG>NNNNNNNNN

Into a 0.5 mL vial place (all on ice while thawing, and while mixing):

DNA vol to get 30,000 ng DNA
20 uL of 10x SAU buffer
2 uL of BSA (comes with the enzyme)
12 uL of SAU enzyme
Water to a total of 200 uL

[A half reaction can be done if less than 30,000 ng of DNA is available.]

Put in the enzyme LAST and do not vortex, just gently add the enzyme up-and-down with the pipette.

We use for all our work **Sigma** water (**#W3500**). There are many sources for "PCR grade water" which is very expensive. The Sigma water has never failed us.

Wrap the top of the vial with Parafilm, and then incubate in a water bath set at 37° C overnight.

The next day, run out a few mL of the mix on a 1% agarose gel to see if digestion is complete. If any high molecular weight DNA remains, then add more enzyme and incubate for several more hours.

Once done, heat the vial to 65 ° C for 20 min to kill the enzyme.

STEP 3. Selection for DNA fragment size.

There is very little parasite DNA in the digested mix, so care must be taken to prevent any damage to the DNA at this point.

Prepare a 1% agarose gel and gel rig -- clean the entire rig with plenty of water, then add fresh buffer. It is important to clean out any residual ethidium bromide or other stain that may be in the rig.

DO NOT ADD ETHIDIUM BROMIDE TO THE GEL AT THIS POINT!

Add 10 uL of loading dye directly into the mix (or less if a half reaction). First test with a duplicate gel to find the maximum amount of water+dye that can fit into a well on the rig. For ours, it was 22 uL per lane.

Now put a 100 bp ladder on the two outer lanes of the gel, and add the maximum amount of the digested DNA mix per lane. Run the gel (we use 100 mV for 1:10 hrs.).

Remove the gel, and cut off the two ladder lanes. Stain these in ethidium bromide. Cut a piece of aluminum foil that exactly matches the size of the central piece of the gel (the piece with the target DNA). Put all three pieces on the UV illuminator. Fit the two slivers with the ladder back onto the central piece of gel. (We deliberately cut the two slivers in a "sloppy" line to be sure the fit is precise.) Now the gel is back together with only the ladder lanes stained and the rest protected from the UV light.

Using a sterile razor blade, cut out the portion of the gel with the desired DNA size fragments. We divide the gel into three size ranges:

High molecular weight = > 4000 bp
Mid molecular weight = 1200 - 4000 bp
Low molecular weight = 400 - 1200 bp

We chose to enrich the low molecular weight fragments.

Stain the remaining portion of the central gel piece in ethidium bromide, then check on the UV box to be sure the right sizes were removed.

The DNA is now removed and cleaned from the gel. We use the Zymoclean Gel Recovery Kit (**Zymo Research, Orange, CA**), and followed the protocol. It is important to follow the instructions, especially for temperatures, exactly. Be sure the gel is completely melted in the first step! Elute each final tube with 8 uL of water, then do a second elution with 8 uL of water. There may be

as many as a dozen tubes to process the entire piece of gel. Combine the first elutions, and combine the second elutions, so will have two vials: Elution I and Elution II. We just used the Elution I vial, but saved the second in case we needed more DNA.

Despite the promises in the Zymo literature, only about 30% of the DNA on the gel was recovered, but the method was very quick and easy.

STEP 4. Placing a linker on the fragments.

The Sau cutter results in an overhang in the DNA fragments (see above). The overhang is GATC on one strand, and CTAG at the end of the other strand.

A linker is added to each end of the DNA with a segment of ds DNA that matches that overhang. This segment is called a linker, and in this case, we call it the SAU-Linker.

Designing the linker is a critical step. The PCR program to be used for malaria DNA requires a very low extension temperature, so the annealing temperature required (based on the linker) must also be low. Therefore, this is the linker that we used:

SAU-A-Linker

5' -- GAT ATA CTA ATG AAT GTT GG -- 3'

SAU-B-Linker

5' -- GATC CCA ACA TTC ATT AGT ATA TC -- 3'

Our oligos were from **Qiagen Operon (Alameda, CA)**.

To make the ds linker prepare:

100 mM SAU-A
100 mM SAU-B
5 M NaCl

Dilute the NaCl: 2uL of 5 M NaCl + 18 uL of water

Mix in a 0.5 uL vial
10 uL of SAU-A
10 uL of SAU-B
2 uL of diluted NaCl

Place the vial into a PCR cycler with this program:

95° C for 2 sec, the ramp down to 21° C over a 25 min period.

Store this linker in the frig.

The linker will thus look like this:

GATC CCA ACA TTC ATT AGT ATA TC
GGT TGT AAG TAA TCA TAT AG

with the GATC overhang that matches the overhang in all the DNA fragments (at both ends).

STEP 5. Ligation of the linker onto all the fragments.

To amplify all the fragments by PCR, a primer site is required. The ds linker is added to all the fragments, thus providing a known primer site. The SAU-A oligo is the only primer needed. Note that it will anneal to one strand at one end and the complementary strand at the other. Thus, only a single primer is needed.

Ligation always has a bit of mystery to it. Ligation that works in one lab for a particular set of conditions, will fail in another lab. Some labs even have their special ligation area on one bench that is their lucky spot. So, try different conditions to determine the best results.

We used the T4 ligase kit from **NEB**.

Keep the enzyme on ice while thawing! On ice add together:

DNA 7 uL
Linker 3 uL
Ligase buffer 2 uL
Water 7 uL
Mix gently
Gently add 1 uL of the ligase
This gives a total of 20 uL

Tabletop, and room temperature for 35 minutes worked for us (on our lucky spot on a single bench!)

Heat at 70° C to kill the enzyme.

STEP 6. Amplify the DNA.

Here is one of the critical steps that makes this procedure work for the parasite DNA.

Plasmodium DNA is highly AT-rich. Stretches of DNA with a high AT content denature at a low temperature. The 72° C extension step that is normal for PCR (the taq enzyme functions best at that temperature) will actually melt AT-rich DNA. Thus, the DNA is denaturing as fast as it is trying to extend. Thus, the normal PCR extension temperature will favor the host DNA at the expense of the parasite DNA! The solution is the use a low extension temperature, perhaps 60° C. The extension time must be increased. This is why the linker needs to be designed to provide a low annealing temperature.

Next, to FAVOR the parasite DNA, use a low melting temperature. We used 84° with success, but an even lower temperature may work even better, perhaps 80°. At these low temperatures, the host DNA will denature only very slowly, but the parasite DNA will melt.

The first step is to heal any nicks that remain in the ligated DNA+linker. This step is done only for this first PCR.

PCR mix (25 mL reaction in a 0.5 mL vial)

2 uL linked DNA
3 uL 10 uM SAU-A
20 uL water

put 25 uL of this into a vial with a Ready-to-Go PCR bead (**Amersham**). This system has an optimal mix of taq enzyme, dNTP's, and buffers. After trying many prepared mixes, and making our own, we always come back to the beads. We get the best results by far using the Ready-to-Go beads. They are expensive, and the price is rising, but they save a great deal of time and misery.

The healing step is 60° for 1 hr.

Followed by the program:

84° C for 2 min
32x
84° for 1 min
52° C for 10 sec
48° C for 10 sec
60° C for 4 min
then a final extension at 60° for 4 min

Run out on a gel to be sure that amplification has taken place (this tests if the linker ligated successfully).

STEP 7. Picking the probe.

The goal during the enrichment is to pick out the fragments containing a microsatellite. The PCR product from the above step will have favored the parasite DNA, but lots of host DNA was also amped. Thus, the specific microsatellite chosen for the search should also favor the parasite.

Plasmodium DNA is AT-rich, so microsatellites most likely will also be AT-rich.

For reading the genotype results from an ABI instrument, triplet repeats (3-base) are easier to read than 2-base. Thus, we decided to enrich for a 3-base repeat that was AT-rich.

A visit to the BLAST site confirmed that CG-repeats are rare in the *Plasmodium* genome. But, the ATT repeat is common, but relatively rare in a vertebrate genome. *Mus*, human, and *Gallus* full genomes are available, and a searched that the ATT repeat is at least 30x more common in the *Plasmodium* genome.

The ATT x 10 repeat was chosen (30 bp long).

This probe was ordered from **Qiagen Operon**, with a biotin molecule added to the 5' end. Kits are available to add biotin to oligos, but if only one or a few probes are needed, it is cost (and time!) effective to order the oligo with biotin already added.

Biotin is a B-vitamin ($C_{10}H_{16}O_3N_2S$)

STEP 8. Annealing the probe to the target microsatellite.

A hybrid solution is prepared:

300 uL of 20X SSC
10 uL of 10% SDS
690 uL of water

All of these reagents were ordered from **Sigma**.

Hybrid reaction:

40 uL of PCR product from STEP 6 (several duplicate PCR's were run)
4 uL of 10 uM biotin-labeled ATT probe

56 uL of the hybrid solution

Heat to 95° C for 15 min, then into a 45° water bath (wrap the vial in parafilm).

Incubate in the waterbath for 18 hours (overnight with careful planning).

STEP 9. Prepare reagents for capture of fragments.

Need:

TBST =
100 mM tris at pH 7.5
150 mM NaCl
0.1% Tween 20

These are a bit tough to mix. To make 1 L (needed for autoclaving), mix 12.11 g Tris base (NOT Tris HCl) in 700 mL of water. Add concentrated HCl until get pH of 7.5. Not all pH probes work with Tris buffers! Be careful to use the correct probe. This will take about 15 mL to get to a ph of 8.0. Add 8.766 g of NaCl. Add 1 mL of Tween-20. Then, add water ... testing pH and adding more HCl if needed. Add water to 1 L. Test pH again. Autoclave at 20 minutes. can make a half mix for 500 mL.

TBT=
Same as TBST without the NaCl

SSC/SDS=
0.2% SSC
0.1% SDS

Mix by using 20X SSC and 10% SDS solution (**Sigma**). 98 mL of water+1 mL of SDS+ 1 mL of SSC

TE (a TE with low EDTA)=
10 mM Tris (NOT tris HCl)
0.1 mM EDTA

Mix by using 0.5 M EDTA solution
25 uL EDTA and 0.151 g Tris in 126 mL of water and autoclave.

STEP 10. Capture of fragments.

Need Avidin D beads from **Vector Labs (Burlingame, CA**). Avidin has a very high affinity for biotin.

Prepare the beads. Put the vial containing the beads (kept in the frig after receiving from Vector) onto the vortex with foam head to dance the vial around on the top. Takes quite a while for the beads to go into a slurry.

Take 100 mL of the slurry (only approximate)
Spin down (can do for 20 sec at 12,000, but this will make the beads tough to get back into a slurry, or under slow speed, but they won't be as clean).
Remove supernatant throw away
Add 400 uL of TBST at 45° C (keep the mixes in a bath or incubator set at 45° C for about an hour before using).
Vortex gently
Spin down and toss out the supernatant
Wash again with TBST, spin, toss

Add 300 uL of TBT to the DNA that has been incubating overnight
Mix with pipette and stir a bit
Put some of this into the beads to get the bead back up into the liquid
Put beads and DNA into a 1.5 mL vial
Incubate with shaking (with vortexer with foam on top) at 45° C for 30 minutes or more. We simply put a vortexer into an incubator set at 45° for an hour or so before using, then shake within the incubator.

Spin down, take off supernatant (can be saved for troubleshooting)
Wash the beads 4 times with 800 uL of TBST at 45° C
then
Wash the beads 4 times with 800 mL of SSC / SDS at 45° C

Again, if use the fast spin, will get cleaner beads, but tough to get them back into a slurry. Therefore, we use a short slow spin, and just wash many times as described above. Each time, be sure to get as much of the supernatant as possible off the beads.

After last spin, add 100 uL of the TE
Mix
Put into a 0.5 mL vial
Put into a PCR cycler for at 95° C for 5 minutes
Flick the vial every 2 minutes
Spin down, take off and SAVE the supernatant.

Clean up with a **Qiagen** or **Sigma** PCR clean-up kit.

What has happened? When the DNA was heated to 95°, it denatured into ssDNA. During the hybrid incubation, conditions were set to favor annealing of the biotin-labeled probe onto complementary DNA sequence to the ATT repeat (thus the 45° temperature). The mix now has lots of the loose probe, all kinds of ssDNA and some dsDNA, plus segments with the ATT repeat

with the probe attached. When the mix is added to the avidin beads, the biotin attaches to the beads, carrying the attached ATT repeats on DNA fragments. Spinning down the beads and washing discards all the other DNA fragments. Heating the cleaned beads releases the biotin-labeled probe (the ds DNA denatures into ssDNA). Spin down beads again, then taking off the final supernatent will yield DNA fragments with the repeat. Also, other DNA will be stuck to the beads, even after washing 8 times, but we have favored the fragments with the repeat.

STEP 11. PCR enriched DNA.

Use the PCR program above and

3 uL of SAU-A
2 uL of the DNA from above step
20 uL of water
Ready-to-Go bead

Do several replicates to produce DNA for next enrichment.

Run out on a gel to see the product (a smear on the gel).

STEP 12. Two more enrichments.

With careful planning, an enrichment can be done in a day (with overnight incubation). We found by doing 3 enrichments, and cleaning beads very well, we had a very high proportion of cloned fragments that contained the repeat (86%). Thus, no further probes are needed (no blots). A way to be sure cloned DNA contains a fragment is described below.

After the last enrichment, the PCR program was altered to allow a poly A tail to be added for cloning. We used a 20 minute final extension step.

STEP 13. Cloning

We use the TOPO cloning kit from **Invitrogen**. This system is very easy to use, and does not require the blue/white selection of colonies. Any bacteria that do not contain a vector with an insert do not live, and thus all of the colonies seen on the plate contain an insert. We followed the instructions that are provided with the kit. When we first started this work we made two mistakes: First, we were not careful to keep our cells at -80 before use, and had poor results (-72 just doesn't keep the cells alive). Second, preparing plates is a bit of an art, and ours were too wet and frothy. An experienced tech was consulted for lessons in how to make good plates.

Colonies were picked and put into 20 uL of water, then heated at 95° C for 10 min. We call these "colony boils" and no growing up of the colony in liquid media was done. Thus, we had only had a few shots to amplify each colony. We found this cheaper and faster than growing bacteria and recovering DNA from the liquid culture. Colony boils were stored in the frig.

We used the TOPO primers (ordering more from Operon, rather than the expensive sets from Invitrogen).

PCR of each colony was

1 uL of T3 primer
1 uL of T7 primer
8 uL of colony boil DNA (not purified, just the bacteria mixed with water and after heating to break open the bacteria).
15 uL of water
Ready-to-go bead

Program used was

94° C for 2 min + 32x (94° C / 1 min, 56° C / 1min, 60° C / 3 min) + final extension at 60 / 15 min.

Run out on a gel to see which colonies contained a fragment of a useful size.

STEP 14. Saving time with sequencing.

We sequenced 156 clones. As noted above, about 14% did not contain a repeat. Also, for *P. mexicanum*, we got mostly very long repeats, some over 200x (600 bp!). Such long repeats are found in the *P. falciparum* genome, but are rare. Why *P. mexicanum* has so many is unknown. So, we sought a method to reduce the number of useless sequences that were run.

In the PCR from the colony boils, add a third primer. This is the ATT x 10 repeat (WITHOUT the biotin label). Clones that did not contain a repeat produced a clean, sharp band on the gel. Clones with a long repeat produced a long smear because the third primer sat down at many places on the repeat. Clones with a short repeat produced a short smear. Thus, by adding the third primer, it is possible to select the clones with a short repeat.

Next, do another PCR with just the two primers for the clones that have the short repeat. Doing the second PCR is faster and cheaper than doing all the extra sequences.

STEP 15. Sequencing.

We did sequencing with the M13 forward primer supplied with the cloning kit. Sequencing was done on an ABI instrument.

STEP 16. Identifying parasite microsatellites.

Flanking regions to the repeat were BLAST searched to find those that appeared to be parasite. This was a poor method. Designing primers and testing with infected vs. noninfected lizards was the best way to identify parasite loci. Because of the AT-rich parasite genome, designing primers was difficult. A gradient thermal cycler was a big help in this work.

STEP 17. Multiplexing

As noted above, the high AT content of the parasite genome made design of primers difficult. We could not get several pairs of primers that amped with the same PCR program, and therefore we did not multiplex in the PCR mix. With PCR for microsatellites of "normal" organisms, it is possible to run ten or even more pairs of primers in one PCR with the same program. Alas, with *Plasmodium mexicanum*, we needed to do each locus in its own PCR run. However, we did multiplex the PCR product when running on the ABI instrument (with each locus's product with a different dye on the forward primer).

NOTE: Finding microsatellites in other species.

The above methods can be slightly modified to find microsatellites for other organisms. The linker must be modified for a best annealing temperature in the PCR. Also, the most common microsatellite for that species must be determined (BLAST search for similar species with whole genome sequencing done would be the way to start). With a different repeat probe, the conditions in the hybrid step must be altered to favor that particular probe.

Most repeats that are not completely AT would use a 50° incubation temperature for overnight. The washes would also be at 50° or even 55° C. The melt temperature for the probe+DNA can be calculated using one of the web-based primer design programs (Operon has a good one).

Here are SAU linkers that can be used for other species (suggested by Guiyun Yan of University of California - Irvine):

A = 5' GATC CCA AGC TTC CCG GGT ACC GC 3'
B = 5' GGT TCG AAG GGC CCA TGG CG 5'