CHAPTER 2

Materials and methods

2.1 Study population and sample selection

A retrospective study was conducted using plasma and data collected during two large perinatal HIV prevention trials conducted in Thailand, named PHPT-1 and PHPT-2 [148, 149]. In a previous study, a total of 3,345 HIV-infected pregnant women were tested for HBsAg using an enzyme-linked immunosorbent assay (ELISA, ETI-MAK, DiaSorin, Italy) and 245 were found HBsAg positive [24]. Among these women, HBeAg was tested using ELISA (ETI-EBK, DiaSorin, Italy) while HBV DNA levels were quantified using the Cobas Amplicor HBV Monitor test (Roche Diagnostic Systems, Branchburg, New Jersey, USA) or Abbott real-time HBV DNA™ assay (Abbott laboratories,USA).

To achieve the full-length HBV sequencing, samples of 56 pregnant women with HBV DNA levels equal to or greater than 3.5 $\log_{10}$ IU/mL were selected for this study. This study was divided into 2 sub-studies (Figure 2.1)

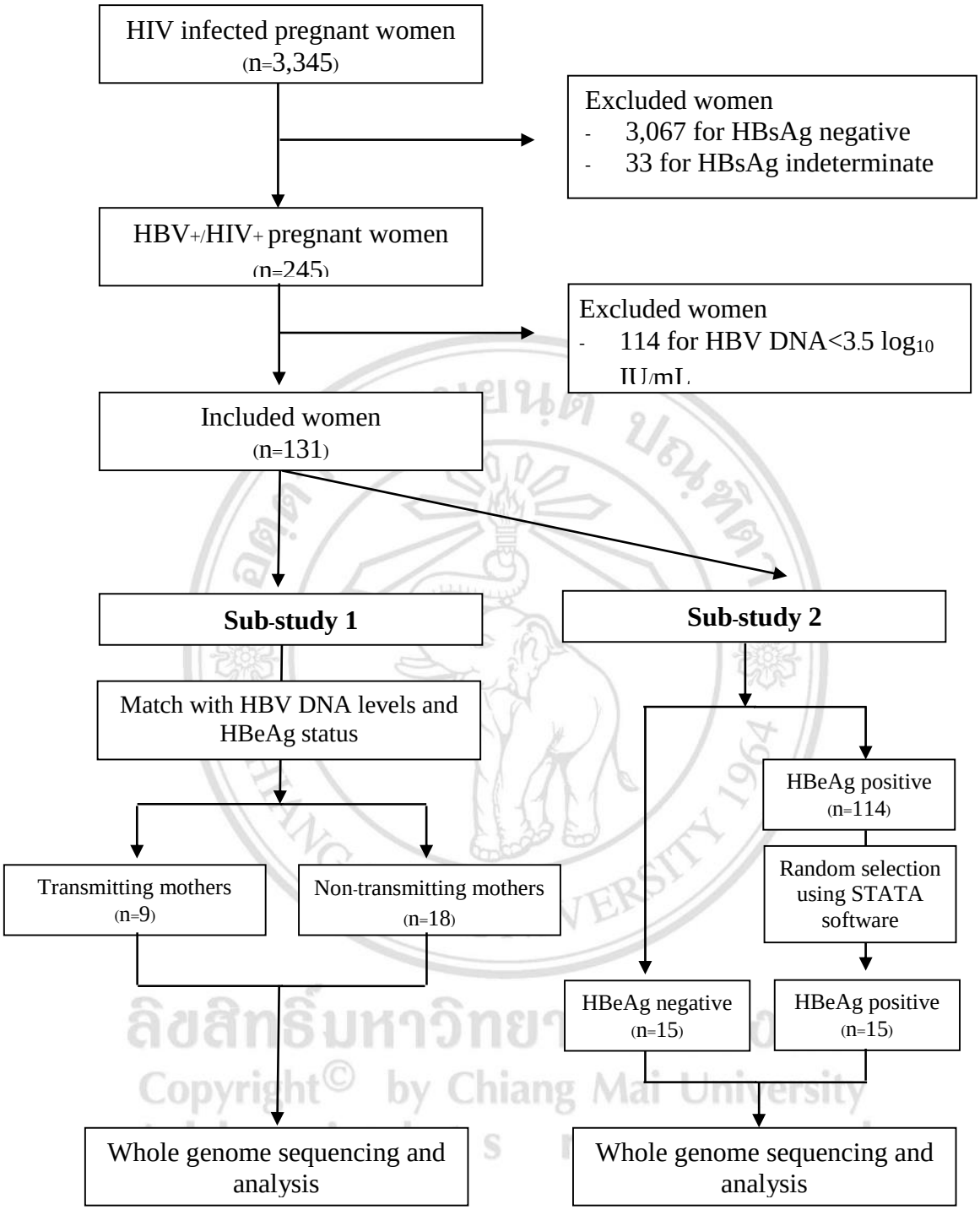


Figure 2.1 Schematic diagram of study design

Sub-study 1: To assess the relationship between HBV genetic diversity and perinatal HBV transmission, HBV transmitting mothers (cases) were matched to HBV non-transmitting mothers (controls) on a ratio of 1:2 according to their HBV DNA levels and HBeAg status.

Sub-study 2: To assess the relationship between HBV genetic diversity and absence of plasma HBeAg in high HBV DNA level, 15 women with HBeAg negative and 15 women with HBeAg positive status were randomly selected for risk factor analysis.

This study was ethically approved by the Ethics committee of the Faculty of Associated Medical Sciences, Chiang Mai University (Reference number: AMSEC-57EM-011). The ethics committee did not consider the patient consent necessary, since the samples used were retrospective and confirmed as HBV-infected individuals by previous study [24].

2.2 HBV DNA extraction

The selected plasma samples were removed from the freezer and allowed to thaw at room temperature. A 1.0 mL of each sample was subjected into the Abbott m2000sp instrument (Abbott Molecular Inc., USA), an automated sample preparation system designed to use magnetic micro-particle processes for the purification of DNA from plasma. DNA extraction was performed following the machine protocol and reagents manufacturer's instructions (https://www.abbottmolecular.com/enus/staticAssets/pdfs/us/realtime-hbv-package-insert.pdf). HBV DNA was extracted by 4 steps including lysed the virions, captured the DNA, washed the particles to remove unbound sample components and eluted the bound DNA and transferred to a 96-deep well plate. Proteinase K enzyme was included in the extraction steps to digest proteins associated with DNA. Finally, the DNA was then ready for amplification.

2.3 PCR amplification of full-length HBV genomes

Full-length HBV DNA was amplified by polymerase chain reaction (PCR) using a primer pair, which has a *Sap*I restriction site, introduced by Gunther *et al.* [150]. The PCR was performed in a 50 µl volume using Platinum PCR High Fidelity (Invitrogen, Carlsbad, CA) and specific primers: P1-F and P2-R [150, 151] (Figure 2.2), which

provide PCR fragments of full-length HBV genome (3.2 kb). The cycling conditions was initiated by denaturation at 94°C for 2 min, followed by 40 cycles of denaturation at 94°C for 40 sec, annealing at 60°C for 1.5 min, and extension at 68°C for 3 min with an increment of 12 second per PCR cycle. The 3.2 kb PCR products were stained with ethidium bromide and visualized by 1% agarose gel electrophoresis.

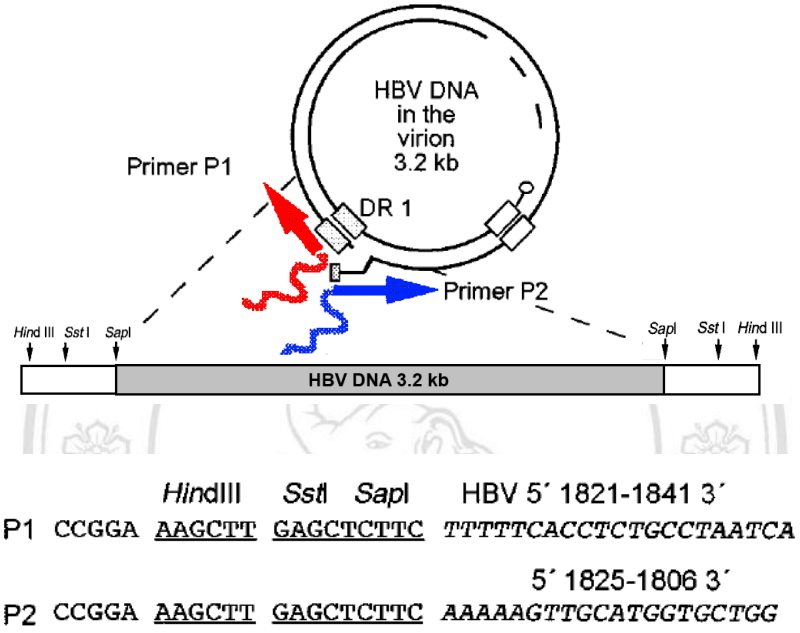


Figure 2.2 Sequences and binding positions of the primer P1 and P2 used for full-length HBV genome amplification [150].

2.4 HBV Surface gene Amplification

Hepatitis B virus surface gene (681 bp) was amplified by single round PCR using a HBV specific primer A3R and A5F (Table 2.1, Figure 2.3), introduced by Gunther *et al.* [150]. The PCR was performed in a 50 µl volume using Platinum PCR High Fidelity (Invitrogen, Carlsbad, CA, USA). The cycling conditions was initiated by denaturation at 94°C for 2 min, followed by 40 cycles of denaturation at 94°C for 40 sec, annealing at 58°C for 30 sec, and extension at 68°C for 2 min. This PCR condition provides PCR fragments of 1,056 bp, which cover the entire surface gene. Then, PCR products were stained with ethidium bromide and visualized by 1% agarose gel electrophoresis.

Table 2.1 HBV-specific primers for surface gene amplification

Primer name	Nucleotide position	Nucleotide sequence	References
A3-R	1121 – 1100	AGA AAG GCC TTG TAA GTT GGC G	[151]
A5-F	66 – 90	GCT CCA GTT CAG GAA CAG TAA ACC C	[151]

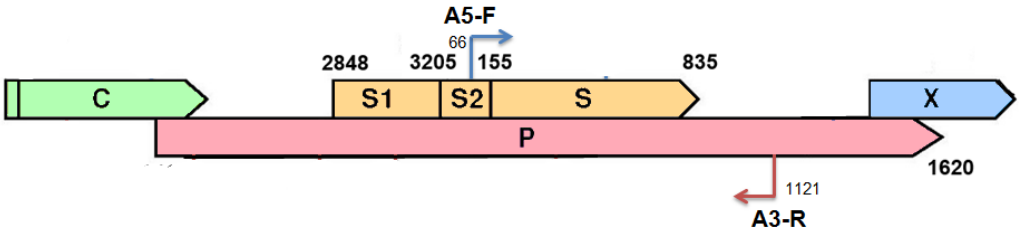


Figure 2.3 Binding positions of primers used for HBV Surface gene amplification

2.5 Purification of PCR Product

The PCR containing 3.2 kb products or 1,056 bp was purified using E.Z.N.A. Cycle Pure Kit (Omega Bio-tek, GA, USA) following the manufacturer's instructions. Briefly, 45 microliters (µL) of each PCR product was thoroughly mixed with 220 µL CP Buffer and centrifuged to collect any drops from the inside of the lid. Then, the mixture was load to a HiBind® DNA Mini Column with collection tube and centrifuged at 13,000x g for 1 minute at room temperature. DNA wash buffer were used to wash each sample two times. This step was followed by centrifuge the HiBind® DNA Mini Column at 13,000x g for 2 minutes to dry the column. Elution of the final purified PCR products was accomplished by directly add 50 µL of elution buffer to the center of column matrix and centrifuged at 13,000x g for 1 minute. Prior to sequencing, purified PCR products were verified again with 1% agarose gel electrophoresis.

2.6 Direct sequencing of HBV full-genomes

The whole genome of HBV was sequenced with a set of HBV-specific primers (Table 2.2, Figure 2.4), using the Dye Terminator Cycle Sequencing Kit (BigDye Terminator Mix V. 3.1, Applied Biosystems, USA). The sequencing products were analyzed on an ABI 3100 genetic analyzer (Applied Biosystems, USA).

Table 2.2 A set of HBV-specific primers for HBV whole genome sequencing.

Primer name	Nucleotide position	Nucleotide sequence	References
A1-F	2357 – 2380	GGC AGG TCC CCT AGA AGA AGA ACT	[150]
A2-R	477 – 455	GGA CAA ACG GGC AAC ATA CCT TG	[151]
A3-R	1121 – 1100	AGA AAG GCC TTG TAA GTT GGC G	[151]
A4-F	676 – 699	TTT ACT AGT GCC ATT TGT TCA GTG	[151]
A5-F	66 – 90	GCT CCA GTT CAG GAA CAG TAA ACC C	[151]
A6-R	2432 – 2408	ATT GAG ATC TTC TGC GAC GCG GCG A	[151]
A10-F	1266 – 1286	CCA TAC TGC GGA ACT CCT AGC	[150]
A13-R	2957 – 2935	TTG GGA TTG AAG TCC CAA TCT GG	[150]

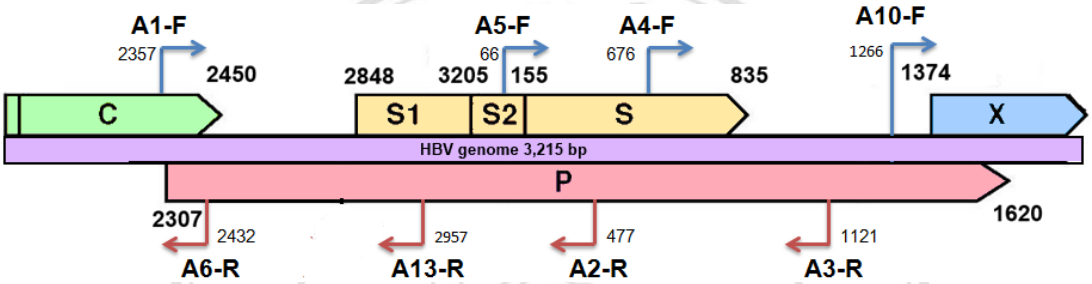


Figure 2.4 Nucleotide positions and direction of primers binding used for direct sequencing of HBV full-genomes

2.7 Sequence analysis

Sequence analysis was performed with BioEdit software (www.mbio.ncsu.edu/bioedit). The alignment of the full-length sequence was performed using the ClustalW method (www.ebi.ac.uk/Tools/msa/clustalw2) and the phylogenetic tree was constructed on HBV reference sequences and sample sequences by a neighbor joining tree using Molecular Evolutionary Genetics Analysis (MEGA 6.0) software (Biodesign Institute, Tempe, AZ, USA).

2.8 HBV genotyping

The HBV genotype was determined by Phylogenetic tree analysis using MEGA6.0 software. The phylogenetic tree analysis was constructed using neighbour-joining method, and based on full-length HBV reference sequences (genotype A–J). Reference sequences obtained from GenBank database under the following accession numbers: genotype A (X70185, V00866, X51970, M57663, X02763, S50225, AP007263), B (D00331, D00329, M54923, AB540582, DQ361535, D00330), C (X01587, M38454, AB540585, V00867, AF068756, EU498227, M12906), D (X59795,

M32138, AB188243, X02496), E (X75664, X75657, AP007262), F (X75663, X75658, X69798, X75658), G (AF160501), H (AB516395, EU498228), I (FJ023659, FJ023672) and J (AB486012). The sequence of the woolly monkey HBV (WMHBV, accession number NC028129, AY226578 and AF046996) was used as an out-group to root the phylogenetic tree. HBV sample sequences were grouped based on phylogenetic tree. The S genes obtained from these 37 full-length genome sequences were then analyzed separately using the same method to identified HBV genotype of sample with only the *surface* (*S*) region available.

HBV genotype was further confirmed by two web-based tools: 1) the HIV-GRADE HBV tool (www.hiv-grade.de/cms/grade/explanations/hbv-tool/), which is a web-based genotyping analysis of the *surface* and *reverse transcriptase* (*RT*) genes, and 2) the NCBI Genotyping tool, a web-based genotyping tool from the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/projects/genotyping/help.html), which investigates the genotype by analysis of the full-length genome sequences.

2.9 Genetic diversity analysis

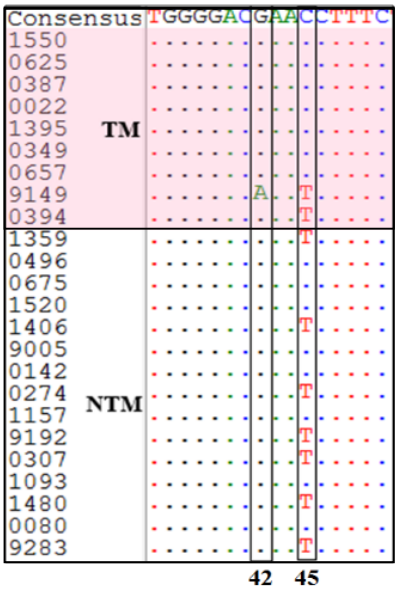
2.9.1 Shannon entropy score

The Shannon entropy score of each nucleotide or amino acid position was calculated according to the following equation where p_i is the probability of nucleotide or amino acid *i* occurring at the position and *n* is the total number of sample sequences per group.

$$H_n = -\sum_{i=1}^{n} p_i \ log_b \ p_i$$

An invariant position has a Shannon entropy score of zero (Figure 2.5). The highest Shannon entropy score is dependent on the number of nucleotide or amino acid in each position of sequence. The Shannon entropy score was used as a way to assign a score to each nucleotide or amino acid that reflects the variability in that position. This method is potentially useful for assess the diversity in the population in a cross-sectional study. For biological purpose and understanding selective pressure on a site a phylogenetic method,

comparative measures of synonymous and non- synonymous substitution rates would be preferred.



Position	Consensus	TM entropy	NTM entropy
40	A	0	0
41	C	0	0
42	G	0.349	0
43	A	0	0
44	A	0	0
45	C	0.530	0.691

Figure 2.5 Example results for Shannon entropy score of two populations (TM and NTM). At nucleotide position 42, TM group has 89% G and 11% A and Shannon entropy score of 0.349 that higher than the score of NTM group which has no variant (Shannon entropy score of zero).

2.9.2 Genetic distance analysis

Genetic distance (GD) is a measure of the degree of gene difference between populations. GD were calculated by counting the number of sites which differ between two sequences divided by the sequence length using the Kimura 2-parameter model [152] conducted in Molecular Evolutionary Genetics Analysis version 6.0 (MEGA6) [153].

2.9.3 The dN/dS ratio

Genetic diversity was determined by the number of synonymous substitutions per synonymous site (dS) and the number of non-synonymous substitutions per non-synonymous site (dN) using the Tamura-Nei model [154] as implemented in MEGA6 software[153]. These dN/dS ratio give an estimate selection; positive selection (positive constraint) is inferred if dN/dS is greater than 1; negative selection (conservative constraint) is inferred if it is less than 1.

2.10 Statistical analysis

Baseline characteristics of study population, including maternal age, HBV DNA levels, ALT levels, CD4 cell count and HIV viral load were described using percentage for categorical data and median with interquartile range (IQR) for continuous data. Continuous variables, including maternal age, HBV DNA levels, ALT levels, CD4 cell count and HIV viral load were transformed into categorical by their common cut-off values or their median values.

Sub-study 1: Genetic diversity was tested for association with HBV perinatal transmission using Fisher's exact tests.

Sub-study 2: Genetic diversity was tested for association with the absence of HBeAg using univariate exact Fisher's exact tests.

The differences were considered statistically significant if *p*-values less than 0.05.