REFERENCES

Alexander, D.H., Novembre, J. and Lange, K., 2009, "Fast model-based estimation of ancestry in unrelated individuals", **Genome Research**, Vol. 19, No. 9, pp. 1655-1664.

Allard, M.M., Wilson, R., Keith, L.M. and Budowle, B., 2004, "Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set", **Legal Medicine**, Vol. 6, No. 1, pp. 11-24.

Ansorge, W.J., 2009, "Next-generation DNA sequencing techniques", **New Biotechnology**, Vol. 25, No. 4, pp. 195-203.

Ardlie, K.G., Kruglyak, L. and Seielstad, M., 2002, "Patterns of linkage disequilibrium in the human genome", **Nature Reviews Genetics**, Vol. 3, No. 4, pp. 299-309.

Ballard, J. William O., and David M. Rand. 2005, "The population biology of mitochondrial DNA and its phylogenetic implications", **Annual Review of Ecology, Evolution, and Systematics**, Vol. 36, No. 1, pp. 621-642.

Ballard, J.W.O. and Whitlock, M.C., 2004, "The incomplete natural history of mitochondria", **Molecular Ecology**, Vol. 13, No. 4, pp. 729-744.

Bamshad, M., Wooding, S., Salisbury, B.A. and Stephens, J.C., 2004, "Deconstructing the relationship between genetics and race", **Nature Reviews Genetics**, Vol. 5, No. 8, pp. 598-609.

Bellwood, P., Fox, J.J., Tryon, D. 1995. **The Austronesians: historical and comparative perspectives**. The Australian National University Press, Canberra.

Besaggio, D., Fuselli, S., Srikummool, M., Kampuansai, J., Castrì, L., Tyler-Smith, C., Seielstad, M., Kangwanpong, D. and Bertorelle, G., 2007, "Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences", **BMC Evolutionary Biology**, Vol. 7, Supp. 2, pp. S12.

Bodner, M., Zimmermann, B., Röck, A., Kloss-Brandstätter, A., Horst, D., Horst, B., Sengchanh, S., Sanguansermsri, T., Horst, J., Kramer, T., Schneider, P.M. and Parson, W., 2011. "Southeast Asian diversity: first insights into the complex mtDNA structure of Laos", **BMC Evolutionary Biology**, Vol.11, No. 1, pp. 49.

Cardon, L., 2003, "Using haplotype blocks to map human complex trait loci", **Trends in Genetics**, Vol. 19, No. 3, pp. 135-140.

Chaubey, G., Metspalu, M., Choi, Y., Magi, R., Romero, I.G., Soares, P., van Oven, M. *et al.*, 2010, "Population Genetic Structure in Indian Austroasiatic Speakers: The Role of Landscape Barriers and Sex-Specific Admixture", **Molecular Biology and Evolution**, Vol. 28, No. 2, pp. 1013-1024.

Collins, F.S, 2003, "The Human Genome Project: Lessons from Large-Scale Biology", **Science**, Vol. 300, No. 5617, pp. 286-290.

Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.F., Rosenberg, N.A. and Pritchard, J.K., 2006, "A worldwide survey of haplotype variation and linkage disequilibrium in the human genome", **Nature Genetics**, Vol. 38, No. 11, pp. 1251-1260.

Dermitzakis, E.T., 2010, Population Genetic Principles and Human Populations. In **Vogel and Motulsky's Human Genetics**, Speicher, M.R., Motulsky, A.G. and Antonarakis, S.E. (Eds.), Heidelberg, Berlin, pp. 487-506

Enfield, N.J., 2003, **Linguistics epidemiology : semantics and grammar of language contact in mainland Southeast Asia**. London; New York: Routledge Curzon.

Enfield, N.J., 2005, "Areal linguistics and Mainland Southeast Asia", **Annual Review of Anthropology**, Vol. 34, No. 1, pp. 181-206.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.B. *et al.*, 2007. "A second generation human haplotype map of over 3.1 million SNPs", **Nature**, Vol. 449, No. 7164, pp. 851-861.

Fucharoen, G., Fucharoen, S. and Horai, S., 2001, "Mitochondrial DNA polymorphisms in Thailand", **Journal of Human Genetics**, Vol. 46, No. 3, pp. 115-125.

Galtier, N., Nabholz, B., GléMin, S. and Hurst, G.D.D., 2009, "Mitochondrial DNA as a marker of molecular diversity: a reappraisal", **Molecular Ecology**, Vol. 18, No. 22, pp. 4541-4550.

Gao, X. and Starmer, J.D., 2008, "AWclust: point-and-click software for non-parametric population structure analysis", **BMC Bioinformatics** 9, No. 1, pp. 77.

Gu, S., Pakstis, A.J., Li, H., Speed, W.C., Kidd, J.R. and Kidd, K.K., 2007, "Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations", **European Journal of Human Genetics**, Vol. 15, No. 3, pp. 302-312.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. and Chee, M.S., 2005, "A genome-wide scalable SNP genotyping assay using microarray technology", **Nature Genetics**, Vol. 37, No. 5, pp. 549-554.

Handley, L., Manica, A., Goudet, J. and Balloux, F., 2007, "Going the distance: human population genetics in a clinal world", **Trends in Genetics**, Vol. 23, No. 9, pp. 432-439.

Hawkins, R.D., Hon, G.C. and Ren, B., 2010, "Next-generation genomics: an integrative approach", **Nature Reviews Genetics**, Vol. 11, No. 7, pp. 476-486.

Hill, C., Soares, P., Mormina, M,. Macaulay, V., Meehan, W., Blackburn, J., Clarke, D. *et al.* 2006. "Phylogeography and ethnogenesis of aboriginal Southeast Asians", **Molecular Biology and Evolution**, Vol. 23, No. 12, pp. 2480-2491.

Holsinger, K.E. and Weir, B.S., 2009, "Genetics in geographically structured populations: defining, estimating and interpreting FST", **Nature Reviews Genetics**, Vol. 10, No. 9, pp. 639-650.

Intarapanich, A., Shaw, P.J., Assawamakin, A., Wangkumhang, P., Ngamphiw, C., Chaichoompu, K., Piriyapongsa, J. and Tongsima, S., 2009, "Iterative pruning PCA improves resolution of highly structured populations", **BMC Bioinformatics**, Vol. 10, No. 1, pp. 382.

Irwin, J.A., Saunier, J.L., Strouss, K.M., Diegoli, T.M., Sturk, K.A., O'Callaghan, J.E., Paintner, C.D., Hohoff, C., Brinkmann, B. and Parsons, T.J, 2007, "Mitochondrial control region sequences from a Vietnamese population sample", **International Journal of Legal Medicine**, Vol. 122, No. 3, pp. 257-259.

Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M, Fung, H.C., Szpiech, Z.A. *et al.*, 2008, "Genotype, haplotype and copy-number variation in worldwide human populations." **Nature**, Vol. 451, No. 7181.

Jobling, M., 2004, **Human evolutionary genetics: origins, peoples and disease**. Garland Science Press, New York.

Joyjinda, Y., 2007, The role of mitochondrial background in the expression of leber hereditary optic neuropathy (LHON), Master of Science Thesis, Biochemistry, Faculty of Science, Mahidol University.

Ke, Y., Su, B., Song, X. and Lu, D., 2001, "African Origin of Modern Humans in East Asia: A Tale of 12,000 Y Chromosomes", **Science**, Vol. 292, No. 5519, pp. 1151-1153.

Keinan, A., Mullikin, J.C. Patterson, N. and Reich, D., 2007, "Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans", **Nature Genetics**, Vol. 39, No.10.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., *et al.*, 2003, "Large-scale genotyping of complex DNA", **Nature Biotechnology**, Vol. 21, No. 10, pp. 1233-1237.

Kirov, G., Nikolov, I., Georgieva, L., Moskvina, V., Owen, M.J. and O'Donovan, M.C., 2006, "Pooled DNA genotyping on Affymetrix SNP genotyping arrays", **BMC Genomics**, Vol. 7, No. 27.

Kong, Q.P., Sun, C., Wang, H.W., Zhao, M., Wang, W.Z., Zhong, L., Hao, X.D., *et al.*, 2010, "Large-Scale mtDNA Screening Reveals a Surprising Matrilineal Complexity in East Asia and Its Implications to the Peopling of the Region", **Molecular Biology and Evolution**, Vol. 28, No. 1, pp. 513-522.

Kumar, V., Langstieh, B.T., Madhavi, K.V., Naidu, V.M., Singh, H.P., Biswas, S., Thangaraj, K., Singh, L. and Reddy, B.M., 2006, "Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes", **PLoS Genetics**, Vol. 2, No. 4, pp. e53.

Kutanan, W., Kampuansai, J., Nakbunlung, S., Lertvicha, P., Seielstad, M.., Bertorelle, G. and Kangwanpong, D., 2011, "Genetic affinity and admixture of northern Thai people along their migration route in northern Thailand: evidence from autosomal STR loci", **Journal of Human Genetics**, Vol. 56, No. 2, pp. 130-137.

Lertrit, P., Poolsuwan, S., Thosarat, R., Sanpachudayan, T., Boonyarit, H., Chinpaisal, C. and Suktitipat, B., 2008, "Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis", **American Journal of Physical Anthropology**, Vol. 137, No. 4, pp. 425-440.

Li, H., Cai, X., Winograd-Cort, E.R., Wen, B., Cheng, X., Qin, Z., Liu, W., *et al.*, 2007, "Mitochondrial DNA diversity and population differentiation in southern East Asia", **American Journal of Physical Anthropology**, Vol. 134, No. 4, pp. 481-488.

Listman, J.B., Malison, R.T., Sanichwankul, K., Ittiwut, C., Mutirangura, A. and Gelernter. J., 2011, "Southeast Asian origins of five Hill Tribe populations and correlation of genetic to linguistic relationships inferred with genome-wide SNP data", **American Journal of Physical Anthropology**, Vol. 144, No. 2, pp. 300-308.

Listman, J.B., Malison, R.T., Sughondhabirom, A., Yang, B.Z., Raaum, R.L., Thavichachart, N., Sanichwankul, K., *et al.*, 2007, "Demographic changes and marker properties affect detection of human population differentiation", **BMC Genetics**, Vol. 8, No. 1, pp. 21.

Liu, Z.J. and Cordes, J.F., 2004, "DNA marker technologies and their applications in aquaculture genetics", **Aquaculture**, Vol. 238, No. 1-4, pp. 1-37.

London, E. 2008. **Thailand condensed : 2000 years of history and culture**. Marshall Cavendish Corp., New York.

Luikart, G., England, P.R., Tallmon, D., Jordan, S. and Taberlet, P., 2003, "The power and promise of population genomics: from genotyping to genome typing", **Nature Reviews Genetics**, Vol. 4, No. 12, pp. 981-994.

Majumder, P.P., 2010, "The Human Genetic History of South Asia", **Current Biology**, Vol. 20, No. 4, pp. R184-R187.

Martinez-Arias, R., Calafell, F., Mateu, E., Comas, D., Andres, A. and Bertranpetit, J., 2001, "Sequence variability of a human pseudogene", **Genome Research**, Vol. 11, pp. 1071-1085.

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P.A. and Hirschhorn, J.N., 2008, "Genome-wide association studies for complex traits: consensus, uncertainty and challenges", **Nature Reviews Genetics**, Vol. 9, No. 5, pp. 356-369.

Melton, T., Clifford, S., Martinson, J., Batzer, M. and Stoneking, M., 1998, "Genetic evidence for the proto-austronesian homeland in asia : MtDNA and nuclear DNA variation in Taiwanese aboriginal tribes", **American Journal of Genetics**, Vol. 63, pp. 1807-1823.

Morin, P., 2004, "SNPs in ecology, evolution and conservation", **Trends in Ecology & Evolution**, Vol. 19, No. 4, pp. 208-216.

Nachman, M.W. and Crowell, S.L., 2000, "Estimate of the mutation rate per nucleotide in humans", **Genetics**, Vol. 156, pp. 297-304.

Ooi, K., 2004, **Southeast Asia a Historical Encyclopedia from Angkor Wat to East Timor**, ABC-CLIO, Santa Barbara.

van Ooij, C., 2009, "Pharmacogenetics: A SNP for hepatitis C treatment failure", **Nature Reviews Genetics**, Vol. 10, No. 11, pp. 738-739.

Oota, H., Pakendorf, B., Weiss, G., von Haeseler, A., Pookajorn, S., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. and Stoneking, M., 2005, "Recent Origin and Cultural Reversion of a Hunter-Gatherer Group", **PLoS Biology**, Vol. 3, No. 3, pp. e71.

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. and Stoneking, M., 2001, "Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence", **Nature Genetics**, Vol. 29, pp. 20-21.

Pakendorf, B. and Stoneking, M., 2005, "Mitochondrial DNA and human evolution", **Annual Review of Genomics and Human Genetics**, Vol. 6, No. 1, pp. 165-183.

Peng, M.S., Quang, H.H., Dang, K.P., Trieu, A.V., Wang, H.W., Yao, Y.G., Kong, Q.P. and Zhang, Y.P., 2010, "Tracing the Austronesian Footprint in Mainland Southeast Asia: A Perspective from Mitochondrial DNA", **Molecular Biology and Evolution**, Vol. 27, No. 10, pp. 2417-2430.

Price, A.L., Zaitlen, N.A., Reich, D. and Patterson, N., 2010, "New approaches to population stratification in genome-wide association studies", **Nature Reviews Genetics**, Vol. 11, No. 7, pp. 459-463.

Reich, D., Thangaraj, K., Patterson, N., Price, A.L. and Singh, L., 2009, "Reconstructing Indian population history", **Nature**, Vol. 461, No. 7263, pp. 489-494.

Rosenberg, N.A, 2002, "Genetic Structure of Human Populations", **Science**, Vol. 298, No. 5602, pp. 2381-2385.

Rosenberg, N.A., Huang, L., Jewett, E.M., Szpiech, Z.A., Jankovic, I. and Boehnke, M., 2010, "Genome-wide association studies in diverse populations", **Nature Reviews Genetics**, Vol. 11, No. 5, pp. 356-366.

Rubinoff, D. and Holland, B., 2005, "Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference", **Systematic Biology**, Vol. 54, No. 6, pp. 952-961.

Scheet, P. and Stephens, M., 2006, "A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase", **The American Journal of Human Genetics**, Vol. 78, No. 4, pp. 629-644.

Shendure, J. and Ji, H., 2008, "Next-generation DNA sequencing", **Nature Biotechnology**, Vol. 26, No. 10, pp. 1135-1145.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K., 2001, "dbSNP: the NCBI database of genetic variation", **Nucleic Acids Research**, Vol. 29, No. 1, pp. 308-311.

Simonson, T.S., Xing, J., Barrett, R., Jerah, E., Loa, P., Zhang, Y., Watkins, W.S., *et al.* 2011. "Ancestry of the Iban Is Predominantly Southeast Asian: Genetic Evidence from Autosomal, Mitochondrial, and Y Chromosomes", **PLoS ONE**, Vol. 6, No. 1, pp. e16338.

Siriboonpiputtana, T., Jomsawat, U., Rinthachai, T., Thanakitgosate, J., Shotivaranon, J., Limsuwanachot, N., Polyorat, P. and Rerkamnuaychoke, B., 2010, "Y-chromosomal STR haplotypes in Central Thai population", **Forensic Science International: Genetics**, Vol. 4, No. 3, pp. e71-e72.

Slatkin, M., 2008, "Linkage disequilibrium: understanding the evolutionary past and mapping the medical future", **Nature Reviews Genetics**, Vol. 9, No. 6, pp. 477-485.

Stoneking, M. and Delfin, F., 2010, "The Human Genetic History of East Asia: Weaving a Complex Tapestry", **Current Biology**, Vol. 20, No. 4, pp. R188-R193.

Stuart-Fox, M., 1998, **The Lao Kingdom of Lan Xang : rise and decline**. White Lotus Press, Bangkok

Stumpf, M.P.H. and Goldstein, D.B., 2003, "Demography, Recombination Hotspot Intensity, and the Block Structure of Linkage Disequilibrium", **Current Biology**, Vol. 13, No. 1, pp. 1-8.

Syvänen, A.C. 2005, "Toward genome-wide SNP genotyping", **Nature Genetics**, Vol. 37, No. 6s, pp. S5-S10.

Tajima, A., Pan, I.H., Fucharoen, G., Fucharoen, S., Matsuo, M., Tokunaga, K., Juji, T., Hayami, M., Omoto, K. and Horai, S., 2002, "Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia", **Human Genetics**, Vol. 110, No. 1, pp. 80-88.

Tamura, K., Dudley, J., Nei, M. and Kumar. S., 2007, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0", **Molecular Biology and Evolution**, Vol. 24, No. 8, pp. 1596-1599.

Tharaphan, P., Chuenkongkaew, W.L., Luangtrakool, K., Sanpachudayan, T., Suktitipat, B., Suphavilai, R., Srisawat, C., Sura, T. and Lertrit, P., 2006, "Mitochondrial DNA haplogroup distribution in pedigrees of Southeast Asian G11778A Leber Hereditary Optic Neuropathy", **Journal of Neuro-Ophthalmology**, Vol. 26, No. 4, pp. 264-267.

The HUGO Pan-Asian SNP Consortium, 2009, "Mapping Human Genetic Diversity in Asia", **Science**, Vol. 326, No. 5959, pp. 1541-1545.

The International HapMap Consortium, 2005, "A haplotype map of the human genome", **Nature**, Vol. 437, No. 7063, pp. 1299-1320.

Vignal, A., Milan, d., san Cristobal, M. and Eggen, A., 2002, "A review on SNP and other types of molecular markers and their use in animal genetics", **Genetics Selection Evolution**, Vol. 34, No. 3, pp. 275-305.

Walley, A.J., Asher, J.E. and Froguel, P., 2009,. "The genetic contribution to non-syndromic human obesity", **Nature Reviews Genetics**, Vol. 10, No. 7, pp. 431-442.

Wang, E.T. 2006. "Global landscape of recent inferred Darwinian selection for Homo sapiens", **Proceedings of the National Academy of Sciences**, Vol. 103, No. 1, pp. 135-140.

Wilder, J.A., Kingan, S.B., Mobasher, Z., Pilkington, M.M. and Hammer, M.F., 2004, "Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males", **Nature Genetics**, Vol. 36, No. 10, pp. 1122-1125.

Wyatt, D. 1986. **Thailand : a short history**. New Haven [Conn.], London.

Xing, J., Watkins, W.S., Shlien, A., Walker, E., Huff, C.D., Witherspoon, D.J., Zhang, Y., Simonson, T.S., Weiss, R.B. and Schiffman, J.D., 2010, "Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping", **Genomics**, Vol. 96, No. 4, pp. 199-210.

Xu, S., Kangwanpong, D., Seielstad, M., Srikummool, M., Kampuansai, J., Jin, L. and the HUGO Pan-Asian SNP Consortium, 2010, "Genetic evidence supports linguistic affinity of Mlabri, a hunter-gatherer group in Thailand", **BMC Genetics**, Vol. 11, No. 1, pp. 18.

Zhang, F., Su, B., Zhang, Y.P. and Jin, L., 2007. "Genetic studies of human diversity in East Asia", **Philosophical Transactions of the Royal Society B: Biological Sciences**, Vol. 362, No. 1482, pp. 987-995.

Zimmermann, B., Bodner, M., Amory, S., Fendt, L., Röck, A., Horst, D., Horst, B., Sanguansermsri, T., Parson, W. and Brandstätter, A., 2009, "Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai)", **International Journal of Legal Medicine**, Vol. 123, No. 6, pp. 495-501.

APPENDIX A: PROTOCOLS FOR GENETIC DATA PREPARATION

A.1 PED files for 2 combined genotype data (used for ADMIXTURE analysis)

1. Prepare matching label files from existing genotype files. If the genetic data must be obtained from database, prepare another file for labels. ADMIXTURE does not incorporate labels but they are needed after the ancestral estimation is finished.
2. Combine 2 separate genetic files or check one separate genetic file against the SNP database. Write the matching (overlapping) SNP in a separate file.
3. Check for strand compatibility. Some SNP genotypes are of the complimentary strand. This would pose problems in encoding. Use one genetic source as the main source and change the genotypes in another file into the same complimentary bases.
4. Write the resulting combined genotype files for future use. This provides flexibility in including or excluding certain populations. Encoded genotype files cannot be used after populations have been included or excluded from the pool of samples.
5. Calculate major and minor alleles in order to assign numeric codes for the genotypes. (1 1 = homozygous wild type, 1 2 = heterozygous variance, 2 2 = homozygous variance, 0 0 = missing data)
6. Transpose genotype matrix as ped files have individuals as rows and SNP genotypes as columns while original genotype files are usually formatted vice versa.
7. Add 6 additional mandatory columns at the front of each row. These six columns will be omitted in ADMIXTURE ancestral estimation.

A.2 Encoded genotype files for PCA

1. Prepare headers into a separate file. Transpose so that the header become a one-column list.
2. Encode the genotype file into PCA format (0 = wild type, 1 = heterozygous variance, 2 = homozygous variance, 0 = missing data).

A.3 Files for AWClust

1. Prepare a header file or use an existing header file.
2. Encode the genotype file into AWClust format (0 = wild type, 1 = heterozygous variance, 2 = homozygous variance, 0 = missing data).
3. Concatenate the encoded genotype with the prepared header.

APPENDIX B: PYTHON CODES

The codes listed in this section are example of the workflow and logics used to construct the codes and prepare data. Real codes are adjusted according to different format of acquired genetic files. For example, some files require changes in header row read (present/absent) or transposing and some others might not need to be queried from database.

B.1 Querying and encoding into ped file format

```
#!/usr/bin/env python
import sys
import MySQLdb

def encode_genotype(gty) :
    alleles = ['A','T','G','C']
    allele_count = {}
    mj_allele = ''
    mj_count  = 0
    mn_allele = ''
    mn_count  = 0
    for allele in alleles :
        count  = gty.count(allele)
        #allele_count[allele] = count
        if count > 0 :
            if count > mj_count :
                mn_count = mj_count
                mn_allele = mj_allele
                mj_count = count
                mj_allele = allele
            else :
                mn_count = count
                mn_allele = allele
    #return major_allele, minor_allele
    encode_gty = gty.replace(mj_allele + mj_allele,
'0').replace(mj_allele + mn_allele,'1').replace(mn_allele +
mj_allele,'1').replace(mn_allele + mn_allele,'2').replace('NN','-1')
    return encode_gty.split(',')

def transpose(matrix) :
    transpose_matrix = []
    for i in xrange(len(matrix[0])) :
        for row in matrix :
            try :
                transpose_matrix[i].append(row[i])
            except :
                transpose_matrix.append([row[i]])
    return transpose_matrix

def query(sql) :
   cs.execute(sql)
   return cs.fetchall()

file_name = sys.argv[1]
ind_file  = sys.argv[2]

ind_dict = {}
fd = open(ind_file)
fd.readline()  #read one line to skip the header
lines = fd.xreadlines()
for line in lines :
```

```
        sp_line = line.strip().split(',')
        ind_dict[sp_line[0]] = sp_line[1].replace('!','').replace('
','_')
    fd.close()

    fd = open(file_name)
    header = fd.readline().strip().split(',')[1:]
    lines  = fd.xreadlines()
    #open database
    db = MySQLdb.connect(user='thaisnp2', passwd='thaisnp2',
db='thaisnp_b129', host='bioinfo.biotec.or.th')
    cs = db.cursor()

    gty = []
    for line in lines :
        # Old gty
        tmp = line.strip().split(',')
        snpid = tmp[0][2:]
        tmp_gty = ','.join(tmp[1:])
        # Query ThaiSNP2 gty
        sql = '''SELECT genotype, allele_A, allele_B
                FROM ThaiSNP2_Genotype
                LEFT JOIN ThaiSNP2_Affymatrix ON
ThaiSNP2_Genotype.rssnp_id = ThaiSNP2_Affymatrix.rssnp_id
                WHERE ThaiSNP2_Genotype.rssnp_id = %d''' % int(snpid)
        result = query(sql)[0]
        thaisnp_gty = result[0].replace('A',
result[1]).replace('B',result[2])
        tmp_gty += ',' + thaisnp_gty
        encode_gty = encode_genotype(tmp_gty)
        gty.append((encode_gty))

    transpose_gty = transpose(gty)

    #print 'Gty : ', len(gty), len(gty[0])
    #print 'Transpose : ', len(transpose_gty), len(transpose_gty[0])
    for tmp in transpose_gty :
        ped = '1 1 0 0 2 1 '
        for i in range(len(tmp)) :
            if tmp[i] == '-1' :
                a = '0 0'
            elif tmp[i] == '0' :
                a = '1 1'
            elif tmp[i] == '1' :
                a = '1 2'
            elif tmp[i] == '2' :
                a = '2 2'
            ped += a + ' '
        print ped.rstrip()
    fd.close()

    ind_content = []
    for ind in header :
        ind_content.append(ind_dict[ind] + '\n')
    for i in range(32) :
        ind_content.append('Thai2\n')
    fd = open('ind_pop.txt','w')
    fd.writelines(ind_content)
    fd.close()
    db.close()
```

B.2 Checking for matching strands with querying

```
#!/usr/bin/env python
import MySQLdb

def query(sql) :
   cs.execute(sql)
   return cs.fetchall()

#open database
db = MySQLdb.connect(user='thaisnp2', passwd='thaisnp2',
db='thaisnp_b129', host='bioinfo.biotec.or.th')
cs = db.cursor()
gty_dict = {}
fd = open('JHS_Genotype.csv')
header = fd.readline()
lines = fd.xreadlines()
for line in lines :
    sp_line = line.strip().split(',')
    snpid   = sp_line[0][2:]
    gty     = ','.join(sp_line[1:])
    gty_dict[snpid] = gty
fd.close()

fd = open("match_sorted.csv")
header = fd.readline()  #read header
lines = fd.xreadlines() #read from the line next to header line

#snpList = []
print header.strip()
for line in lines :
    sp_line = line.strip().split(',')
    snpid   = sp_line[0][2:]
    #snpList.append(snpid)
    allele  = sp_line[3] + '/' + sp_line[4]
    sql = '''select concat(allele_A, '/', allele_B) from
ThaiSNP2_Affymatrix where rssnp_id=%d''' % int(snpid)
    result = query(sql)[0][0]
    if allele <> result :
        if result.find(sp_line[3]) < 0 :
            compl_base = {'A':'T','T':'A','G':'C','C':'G'}
            allele1, allele2  = compl_base[sp_line[3]],
compl_base[sp_line[4]]
            replace_gty = {sp_line[3]+sp_line[3]:allele1+allele1,
sp_line[3]+sp_line[4]:allele1+allele2,
sp_line[4]+sp_line[3]:allele2+allele1,sp_line[4]+sp_line[4]:allele2+al
lele2}
            for gty in replace_gty :
                gty_dict[snpid] = gty_dict[snpid].replace(gty,
replace_gty[gty])
    print 'rs%s,%s' % (snpid, gty_dict[snpid])
fd.close()

db.close()
```

B.3 Checking for matching SNPs from 2 files, changing into compatible strands

```
#get file2 gty = newThai
fd = open(gtyFile2)
```

```
    file2Header = fd.readline()
    lines = fd.xreadlines()
    for line in lines :
        split_line = line.strip().split(',')
        snpID = split_line[0]
        if file1gty_dict.has_key(snpID) == True :
            gty = ','.join(split_line[1:])
            obs2 = countAllele(gty)  #'A,T'
            obs1 = countAllele(file1gty_dict[snpID]) #'T,A'
            file2gty_dict[snpID] = gty
            if obs2 <> obs1 :
                if obs2.find(obs1[0])<0 and obs2.find(obs1[2])<0 :
                    try :
                        allele1 = compl_dict[obs2[0]]
                        allele_a = obs2[0]
                    except :
                        allele1 = '-'
                        allele_a = '-'
                    try :
                        allele2 = compl_dict[obs2[2]]
                        allele_b = obs2[2]
                    except :
                        allele2 = '-'
                        allele_b = '-'
                    replace_dict = {allele_a+allele_a:
allele1+allele1,\
                                    allele_a+allele_b:
allele1+allele2,\
                                    allele_b+allele_a:
allele2+allele1,\
                                    allele_b+allele_b:
allele2+allele2}
                    for key in replace_dict :
                        file2gty_dict[snpID]=
file2gty_dict[snpID].replace(key,replace_dict[key])
                    #print    snpID,    obs1,    obs2,    replace_dict,
file1gty_dict[snpID], file2gty_dict[snpID]
            print     '%s,%s,%s'     %(snpID,     file1gty_dict[snpID],
file2gty_dict[snpID])

            #if len(obs1)>2 or len(obs2)>2 :
            #    print '%s,%s, %s' %(snpID, obs1, obs2)
    fd.close()
```

B.4 Combining overlapping SNP data from 2 genotype files

```
    #!/usr/bin/env python
    import sys

    if __name__=='__main__' :
        thal_file = sys.argv[1]
        depress_file = sys.argv[2]
        fd = open(thal_file)
        thal_header = fd.readline()
        thal_gty  = {}
        lines = fd.xreadlines()
        for line in lines :
            sp_line = line.strip().split(',')
            snp_id  = sp_line[0]
            gty     = ','.join(sp_line[1:])
            thal_gty[snp_id] = gty
```

```
fd.close()
fd = open(depress_file)
depress_header = fd.readline()
depres_gty = {}

# \ is for new line
print '%s,%s' % (thal_header.strip(), \
                 ','.join(depress_header.split(',')[1:]))
lines = fd.xreadlines()
for line in lines :
    sp_line = line.strip().split(',')
    snp_id = sp_line[0]
    if thal_gty.has_key(snp_id) == 1 :
        # matched
        gty = ','.join(sp_line[1:])
        print '%s,%s,%s' % (snp_id, thal_gty[snp_id], gty)
fd.close()
```

CURRICULUM VITAE

NAME Ms. Sattara Hattirat

DATE OF BIRTH 28 September 1979

EDUCATIONAL RECORD

High School	Trium Udom Suksa School, 1996
Bachelor's Degree	Bachelor of Arts (Chinese Language) Chulalongkorn University, 2000 Bachelor of Science (Biology) Ramkhamhaeng University, 2008
Master's Degree	Master of Science (Bioinformatics and Systems Biology) King Mongkut's University of Technology Thonburi, 2011

SCHOLARSHIP Scholarship from Bioinformatics Program, 2009-2011

PUBLICATIONS

Sattara Hattirat, Chumpol Ngamphiw, Anuncai Assawamakin, Jonathan Chan and Sissades Tongsima. "Catalog of Genetic Variations (SNPs and CNVs) and Analysis Tools for Thai Genetic Studies". **Communications in Computer and Information Science** 115. J. H. Chan, Y-S. Ong and S-B. Cho (Eds). Springer (2010)

Pich Tantichukaitikul, Sattara Hattirat, and Jonathan H. Chan, "Parallel Pairwise Sequence Alignment Algorithm Based on Longest Common Subsequence," **Journal of Information Science and Technology** 2010, 1(2), 1-10, Jul-Dec 2010, ISSN:1906-9553

Pich Tantichukaitikul, Sattara Hattirat and Jonathan H. Chan. Parallel LCS Algorithm for Pairwise Sequence Alignment. **3rd National Conference on Information Technology** (2010)

Mongkol Pangpet, Sattara Hattirat and Kansri Boonpragob. "Transplantation of lichen Parmotrema tinctorum on ventilated nets at Khao Yai National Park, Thailand". **34th Science and Technology of Thailand** (2008)

Sattara Hattirat and Kansri Boonpragob. "Different Quantum Efficiency and Photochemistry of the Lichen *Parmotrema tinctorum* in Sunny and Shaded habitats". **33th Science and Technology of Thailand** (2007)

Kansri Boonprakob and Sattara Hattirat (eds.). **Climate Change and Thailand**, Green Peace (2006)

EMPLOYMENT RECORD

Deputy Director
Thai AIDS Treatment Action Group, December 2011 - present

King Mongkut's University of Technology Thonburi
Agreement on Intellectual Property Rights Transfer for Postgraduate Students

Date.....

Name...Ms. Sattara ...Surname...Hattirat...Student Number...52401712...who is a student of King's Mongkut's University of Technology Thonburi (KMUTT) in...Master's Degree... Program...Bioinformatics and Systems Biology... Field of Study... Bioinformatics and Systems Biology... Faculty/School of...Bioresources and Technology and School of Information Technology... Home Address...59/663... Tambon...Samwatawanok... District...Klong Samwa... Province...Bangkok... Postal Code...10510... Country...Thailand...

I, as 'Transferer', hereby transfer the ownership of my thesis copyright to King's Mongkut's University of Technology Thonburi and National Center for Genetic Engineering and Biotechnology, Thailand who has appointed (Dean's name) Assoc. Prof. Narumon Jeyashoke and Assoc. Prof. Dr. Nipon Charoenkitkarn, Dean of School of Bioresources and Technology and Dean of School of Information Technology to be 'Transferees' of copyright ownership under the 'Agreement' as follows.

1. I am the author of the thesis entitled "Genetic analysis and peopling of Thai population" under the supervision of Asst. Prof. Dr. Jonathan Hoyin Chan who is my supervisor, and Dr. Sissades Tongsima who is my co-supervisor, in accordance with the Thai Copyright Act B.E. 2537. The thesis is a part of the curriculum of KMUTT.

2. I hereby transfer the copyright ownership of all my works in the thesis to KMUTT and National Center for Genetic Engineering and Biotechnology, Thailand throughout the copyright protection period in accordance with the Thai Copyright Act B.E. 2537, effective on the approval date of thesis proposal consented by KMUTT.

3. To have the thesis distributed in any form of media, I shall in each and every case stipulate the thesis as the work of KMUTT and National Center for Genetic Engineering and Biotechnology, Thailand.

4. For my own distribution of thesis or the reproduction, adjustment, or distribution of thesis by the third party in accordance with the Thai Copyright Act B.E. 2537 with remuneration in return, I am subject to obtain a prior written permission from KMUTT and National Center for Genetic Engineering and Biotechnology, Thailand.

5. To use any information from my thesis to make an invention or create any intellectual property works within ten (10) years from the date of signing this Agreement, I am subject to obtain prior written permission from KMUTT and National Center for Genetic Engineering and Biotechnology, and KMUTT and National Center for Genetic Engineering and Biotechnology are entitled to have intellectual property rights on such inventions or intellectual property works, including entitling to take royalty from licensing together with the distribution of any benefit deriving partly or wholly from the works in the future, conforming with the Regulation of King Mongkut's Institute of Technology Thonburi Re the Administration of Benefits deriving from Intellectual Property B.E. 2538.

6. If the benefits arise from my thesis or my intellectual property works owned by KMUTT and National Center for Genetic Engineering and Biotechnology, I shall be entitled to gain the benefits according to the allocation rate stated in the Regulation of King Mongkut's Institute of Technology Thonburi *Re* the Administration of Benefits deriving from Intellectual Property B.E. 2538.

Signature ..... Transferer
(Sattara Hattirat)

Signature ..... Transferee
(Assoc. Prof. Narumon Jeyashoke)

Signature ..... Transferee
(Assoc. Prof. Dr. Nipon Charoenkitkarn)

Signature ..... Witness
(Asst. Prof. Dr. Jonathan Hoyin Chan)

Signature ..... Witness
(Asst. Prof. Dr. Marasri Ruengjitchatchawalya)