

**Jméno diplomanta:** Bc. Adam Vonášek

**Garantující katedra:** KKY

**Název diplomové práce:** Návrh algoritmu pro identifikaci řídkých haplotypů u dárců kostní dřeně

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2	Formální a obsahová stránka práce	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	Vhodnost použitých metod	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
5	Správnost získaných výsledků	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
6	Vlastní přínos	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Doplnění hodnocení, připomínky, dotazy:

Diplomová práce řeší problematiku identifikace řídkých haplotypů. Navržený algoritmus je tvořen dvěma částmi - načítáním dat identifikovaných haplotypů v populacích a vlastním identifikačním algoritmem.

Formální úroveň práce je velmi nízká. V práci zcela chybí úvod. Text je často nesrozumitelný, nepřesně formulovaný, autor často mění terminologii, řada věcí je nedostatečně vysvětlena. Reference uvedené v seznamu použité literatury jsou různě formátované a ve dvou případech se stejná reference vyskytuje v seznamu vícekrát. Předložená práce výrazně překračuje rozsah stanovený zadáním. Některé části např. příklady výsledků identifikace by bylo vhodnější zařadit do příloh.

Algoritmus pro načítání populačních dat z webu <http://allelefrequencencies.net> vytváří tři textové soubory pro tři kategorie populací - P1 populace nejbližší české, P2 populace "evropské" a P3 populace české nejvzdálenější mimoevropské. Algoritmus obsahuje chybné vícenásobné načítání některých populací (USA NMDP Caucasian 2x, North Afrika 3x). Nevhodně zvoleným nastavením filtrů navíc zařadí celou populaci P1 také do P2 a část P2 i do P3. Ze stejného důvodu jsou do populace P2 chybně zařazeny populace, které mezi evropské nepatří (např. Germany DKMS - China minority, Russia Sibéria Chukotka Peninsula Eskimo, ...). Dalším problémem tohoto algoritmu je načítání populací (řádově stovky) s daty s low resolution, která nejsou pro identifikaci řídkých haplotypů využitelná. Autorovo tvrzení, že je vždy výhodnější načíst celý soubor než aktualizovat stávající, je spíše úsměvné a dokládá, že algoritmu načítání není vhodně navržen.

Algoritmus identifikace řídkých haplotypů je velmi komplikovaný. Využívá celou řadu funkcí, z nichž některé nejsou dobře popsány a nemají žádné teoretické nebo jinak podložené zdůvodnění a velký počet parametrů, které jsou nastaveny experimentálně na základě verifikačních dat. Algoritmus vychází spíše než z v teoretické části práce uvedených znalostí (vazební nerovnováha, síla vazby, polymorfismus genů, ...) z principu hledání co největší shody na jednotlivých genech haplotypů a zařazením dalších přepočtů a korekcí se snaží co nejvíce přiblížit verifikačnímu řešení. Z výše uvedených důvodů je prakticky vyloučeno lepší parametrizaci tento algoritmus optimalizovat.

Autor v práci uvádí, že identifikace 100 haplotypových párů ve všech třech populacích trvá cca 12 minut. Z důvodu nevhodně zvoleného řešení (textové soubory) a návrhu algoritmu trvá v podstatě stejnou dobu i identifikace pro jeden haplotypový pár.

V kapitole Možnosti vylepšení a rozšíření algoritmu autor na více než dvou stranách textu navrhuje různá vylepšení a úpravy. Jejich realizace by však znamenala kompletní přepracování stávajícího řešení od základu.

# SOUHLASÍ S ORIGINÁLEM

Celkový přínos předložené diplomové práce je prakticky nulový.

Otázky:

- Proč do populačních dat načítáte i data populací s low resolution?
- Proč jste při návrhu zařazení populací do jednotlivých kategorií (P1, P2, ...) nevycházel z v práci uvedené genetické mapy Evropy (Obrázek 3, str. 27)?
- Proč jste nepoužil některou z volně dostupných databází (mySQL, PostgreSQL, ...)?
- Proč jste důsledně neaplikoval a případně hlouběji nerozpracoval znalosti uvedené v teoretické části práce?

Splnění bodů zadání	<input checked="" type="checkbox"/> úplně	<input type="checkbox"/> částečně	<input type="checkbox"/> nesplněno	
Doporučení práce k obhajobě	<input checked="" type="checkbox"/> ano		<input type="checkbox"/> ne	
Celkové hodnocení práce	<input type="checkbox"/> výborně	<input type="checkbox"/> velmi dobře	<input type="checkbox"/> dobře	<input checked="" type="checkbox"/> nevyhověl
Jméno, příjmení, titul oponenta: Miloš, Fetter, Ing.				
Pracoviště oponenta: KKY				

8.9.2017

Datum



Podpis