

به نام خداوند بخشنده مهربان

ژنتیکا

روشها و
پروتکل‌های
ژنتیکی



www.genetica.ir

بسمہ تعالیٰ



Linkage Disequilibrium

استاد:

دکتر محمد رضا نصیری مقدم

مہدی خجستہ کی

بہار ۱۳۹۱

فهرست مطالب:

۳	 مقدمه
۴	 آشنایی با مفهوم LD
۵	 تعریف LD
۷	 عوامل موثر بر مقدار LD
۱۱	 کاربرد های LD
۱۲	 مارکرهای مناسب برای اندازه گیری LD
۱۳	 مقایسه LD در جمعیت های انسانی و دامی
۱۳	 روش های برآورد LD
۱۹	 محاسبه میزان عدم تعادل لینکاژ با حل یک تمرین ساده
۲۰	 نرم افزارهای محاسبه LD
۲۳	 محاسبه LD در نرم افزار R
۳۰	 فهرست منابع

مقدمه:

قانون هاردی واینبرگ:

در جوامع بزرگ به شرط آمیزش تصادفی و عدم دخالت عوامل بر هم زننده تعادل فراوانی های ژنی از یک نسل به نسل بعد ثابت می ماند و تغییر نمی کنند.
عوامل بر هم زننده تعادل ژنتیکی:

۱- مهاجرت

۲- جهش

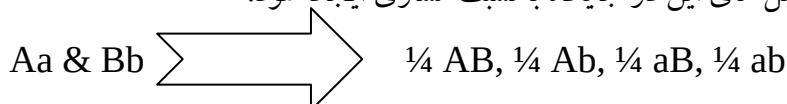
۳- انتخاب

۴- تغییر اندازه موثر جمعیت به سبب Bottleneck, Random drift, Founder effect
Bottleneck: کوچک شدن اندازه جمعیت بواسطه وقوع بلایای طبیعی، جنگ و وقایعی از این قبیل.
Random drift: نمونه گیری تصادفی از آلل ها در جوامع کوچک که فراوانی های ژنی را به سمت حداکثر شدن و یا حداقل شدن پیش می برد.
Founder effect: تاثیر تعداد و نسبت مشارکت مولدین بر شکل گیری یک جمعیت و فراوانی های ژنی موجود در آن.

در شرایطی که قانون هاردی واینبرگ در یک جمعیت برقرار باشد، انتظار میرود که بین فراوانی های هاپلوتایپی و فراوانی های ژنی رابطه ذیل برقرار باشد:

$$p(AB) = p(A) \times p(B)$$

یعنی فراوانی های هاپلوتایپی بصورت مستقیم از روی فراوانی های ژنی قابل محاسبه است. در این صورت اگر دو جایگاه ژنی A و B هر یک با دو آلل بصورت ذیل موجود باشد ممکن است چهار ترکیب هاپلوتایپی از ترکیب آلل های این دو جایگاه با نسبت مساوی ایجاد شود:



نقض قانون هاردی واینبرگ:

در مباحث تئوری قانون هاردی واینبرگ می تواند مصداق داشته باشد اما در دنیای واقعی به سبب وقوع جهش های ناخواسته، تاثیر گرایش تصادفی و حتی انتخاب طبیعی معمولا فراوانی

هابلوتایپ ها نسبت به فراوانی های مورد انتظار متفاوت است. این موضوع نشان دهنده انحراف فراوانی های ژنی از مقادیر مورد انتظار آن ها است.

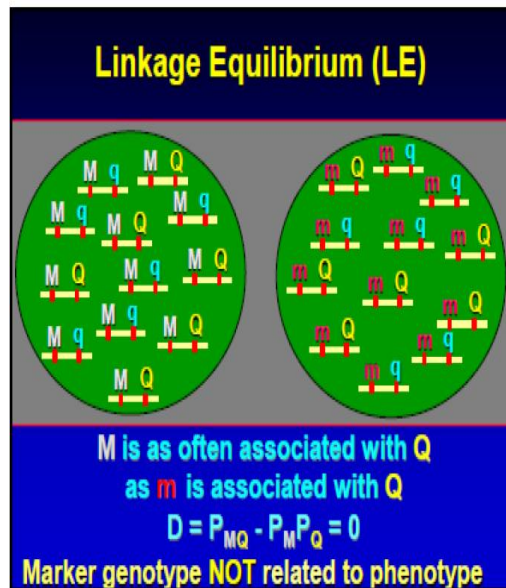
برای تشخیص این انحراف از آزمون k اسکور درباره فراوانی های هابلوتایی موجود و مورد انتظار استفاده می شود. و بر این اساس، فرض صفر آزمون وجود تعادل هاردی واینبرگ در یک جمعیت برای یک جایگاه ژنی است که رد و یا پذیرفته می شود.

آشنایی با مفهوم Linkage Equilibrium در برابر Linkage Disequilibrium:

در بررسی های انجام شده در ژنتیک جمعیت تا کنون فرض بر آن بود که انتقال آلل ها در یک جایگاه ژنی در طول چندین نسل متوالی مستقل از انتقال آلل های مربوط به سایر جایگاه های ژنی انجام می شود. همچنین فرض بر آن بود که قدرت بقا و زیست یک ژنوتیپ در یک جایگاه ژنی توسط جایگاه های دیگر تحت تاثیر قرار نمی گیرد. این دو فرض بصورت مشابه در تعداد قابل توجهی از جایگاه های ژنی نقض شده اند. هنگامی که واریانس ژنتیکی مربوط به دو یا چند جایگاه ژنی بصورت همزمان مد نظر باشد فراوانی های آللی به تنهایی برای توصیف ماهیت پویای جمعیت های طبیعی کافی نیست زیرا مشاهدات با فراوانی آللی و ژنوتیپی مورد انتظار متفاوت است. در این حالت لازم است تا ارتباط غیر تصادفی آلل ها بین جایگاه های ژنی مد نظر قرار بگیرد. این ارتباط غیر تصادفی بین آلل ها در جایگاه های ژنی مختلف بعنوان فاز عدم تعادل گامتی یا بصورت ساده عدم تعادل لینکاژ نام گرفته است. با فرض دو جایگاه M و Q هر یک با دو آلل در صورت برقراری قانون هاردی واینبرگ و وجود تعادل لینکاژ Linkage Equilibrium چهار هابلوتایپ ذیل با نسبت های برابر تولید می شوند. شکل ۱ این موضوع را نشان می دهد:

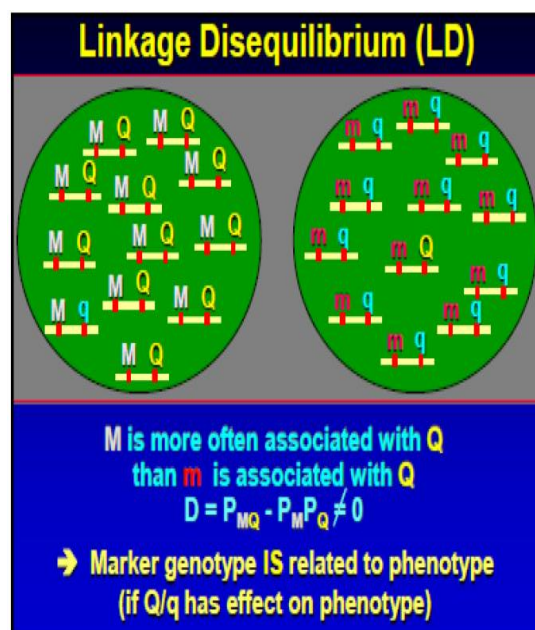
$$MQ \text{ و } Mq \text{ و } mQ \text{ و } mq$$

بنابر آنچه گفته شد تعادل لینکاژ Linkage Equilibrium بین دو یا چند جایگاه ژنی هنگامی مطرح است که ژنوتیپ مربوط به یک جایگاه نسبت به جایگاه دوم مستقل باشد. در این حالت فراوانی های ژنوتیپی و آللی مشاهده در جمعیت با فراوانی های مورد انتظار در قانون هاردی واینبرگ هم خوانی دارد.



$$p(MQ) = p(Mq) = p(mQ) = p(mq)$$

بر همین اساس عدم تعادل لینکاژ بین آلل های دو جایگاه ژنی هنگامی مطرح می شود که ژنوتیپ متعلق به این دو جایگاه نسبت به هم مستقل نبوده و توارث آنها به هم مربوط باشد. در این حالت فراوانی های هاپلوتاییپی از فراوانی های مورد انتظار در قانون هاردی واینبرگ تبعیت نمی کند و برخی هاپلوتایپ ها ممکن است نسبت به مقادیر مورد انتظار در فراوانی بیشتر یا کمتر بسر ببرند.



$$p(MQ) \neq p(Mq)$$

در تصویر بالا هر جا که آلل M وجود دارد به احتمال زیاد آلل Q نیز همراه آن خواهد بود و همچنین است برای آلل های m و q. و این درست همان مفهوم با هم بودن غیر تصادفی آلل ها در کنار یکدیگر است.

در ارتباط با واژه عدم تعادل لینکاژ دو موضوع گمراه کننده است. اول آنکه ارتباط غیر تصادفی بین آلل های دو جایگاه ژنی می تواند اتفاق بیفتد هرچند که دو ژن بهم لینکاژ کروموزومی نداشته باشند. و دوم آنکه به صرف آنکه دو آلل از دو جایگاه دارای لینکاژ کروموزومی باشند این موضوع دلیلی بر وجود عدم تعادل لینکاژ نخواهد بود.

تعریف LD:

بر اساس آنچه در بالا ارائه شد به سبب دخالت عوامل بر هم زننده تعادل، فراوانی های ژنی نسبت به فراوانی های مورد انتظار در قانون هاردی واینبرگ متفاوت خواهد بود و همچنین فراوانی های هاپلوتایپی از فراوانی های ژنی تبعیت نمی کند. بر این اساس هرگونه انحراف فراوانی هاپلوتایپ ها از قانون هاردی واینبرگ را در اصطلاح عدم تعادل لینکاژ یا Linkage disequilibrium می گویند.

تعریف ریاضی LD:

حاصل تفاوت فراوانی مرکب دو آلل از یک مکان ژنی و حاصل ضرب فراوانی این دو آلل ضریب عدم تعادل ژنتیکی نامیده می شود. با در نظر گرفتن دو آلل برای یک مکان ژنی عدم تعادل ژنتیکی با استفاده از رابطه زیر محاسبه می شود:

$$D = P(AB) - P(A).P(B)$$

در این رابطه PAB فراوانی گامت AB، PA و PB به ترتیب فراوانی آللهای A و B می باشند.

تعریف دیگر LD:

با هم بودن غیر تصادفی آللهای مربوط به دو مکان ژنی مختلف سبب ایجاد عدم تعادل لینکاژ می شود. هرگونه پیوستگی بین ژن ها (حتی در صورت عدم پیوستگی فیزیکی آنها بر روی یک کروموزوم) به عدم تعادل لینکاژ مربوط می شود. به این لحاظ پیوستگی ژنتیکی نامتعادل را عدم تعادل گامتی نیز می نامند.

عوامل موثر بر مقدار LD:

عوامل متعددی بر مقدار LD بین دو جایگاه ژنی یا دو مارکر موثر است. بطور کلی هر عاملی که بر تغییر فراوانی های ژنی موثر باشد می تواند بر میزان LD نیز موثر باشد. وقوع جهش ممکن است باعث کاهش یا افزایش LD شود. برای مثال وقوع جهش هایی که باعث حذف یک قطعه کروموزومی می شود ممکن است باعث نزدیک شدن دو جایگاه ژنی به یکدیگر بر روی کروموزوم شده و در نتیجه LD آنها افزایش یابد و یا بعکس جهش هایی که با جایگذاری یک قطعه در بین دو جایگاه ژنی همراه است باعث افزایش فاصله بین دو مکان ژنی و کاهش مقدار LD بین آن ها شود. بطور کلی وقوع جهش های مکرر و با فراوانی بالا با تغییر فراوانی ژنها همراه بوده و در نتیجه مقدار LD را تغییر خواهد داد. عواملی نظیر همخونی، گرایش تصادفی، جریان ژنی (مهاجرت)، انتخاب طبیعی، Founder effect و Population bottleneck با تغییر فراوانی های ژنی باعث افزایش مقدار LD در جمعیت می شود.

جهش:

همانطور که جهش به مقدار اندک باعث تغییر فراوانی های ژنی می شود فرایند جهش نیز باعث ایجاد مقدار زیادی عدم تعادل لینکاژ نمی شود و تنها جهش های مکرر و با فراوانی بالا باعث ایجاد مقداری از عدم تعادل لینکاژ می شوند. هرچند در مناطقی از ژنوم که در آنها نوترکیبی رخ نمی دهد (کروموزوم Y در انسان) جهش می تواند باعث ایجاد عدم تعادل لینکاژ شود که در طول زمان کاهش نیافته بلکه بواسطه گرایش تصادفی و انتخاب افزایش می یابد.

گرایش تصادفی:

در جوامع کوچک با فرض آنکه اندازه موثر جمعیت را با N_e و مقدار نوترکیبی بین دو جایگاه ژنی را با r نمایش دهند مقدار r^2 از رابطه ذیل قابل محاسبه است:

$$E(r^2) \approx 1/[1 + 4N_e C]$$

این رابطه نشان می دهد که اگر $4N_e C$ کوچک باشد مقدار r^2 قابل انتظار برابر با ۱ خواهد بود. برعکس اگر مقدار $4N_e C$ خیلی زیاد شود مقدار مورد انتظار عدم تعادل لینکاژ به ۰ نزدیک می شود. اگر $N_e C$ بزرگ باشد فرمول فوق به رابطه ذیل تبدیل می شود:

$$E(r^2) \approx 1/4N_e \cdot C$$

در روابط فوق $4NeC$ را در اصطلاح نرخ نو ترکیبی جمعیت می گویند. این رابطه از این جهت مهم است که اگر نرخ نو ترکیبی در یک جمعیت مشخص باشد می توان اندازه موثر جمعیت را تعیین نمود.

مهاجرت و جریان ژنی:

این عامل می تواند مقدار قابل توجهی عدم تعادل لینکاژ را در جمعیت ایجاد کند. هر چند این عامل تنها وقتی اتفاق می افتد که فراوانی اللها در دو جایگاه و بین جمعیت های مورد مطالعه متفاوت باشد. هر چه تفاوت فراوانی های ژنی در دو جایگاه و در دو جمعیت بیشتر باشد عدم تعادل لینکاژ ایجاد شده در نتیجه جریان ژنی (مهاجرت) نیز بیشتر خواهد بود. اگر در جمعیت پایه بین دو جایگاه ژنی عدم تعادل لینکاژ وجود داشته باشد این موضوع به شکل گیری و افزایش عدم تعادل لینکاژ بین دو جایگاه در جمعیت مقصد کمک می کند. تقسیم جمعیت اصلی به زیر جمعیت های مختلف باعث تاخیر در میزان کاهش عدم تعادل لینکاژ بین جایگاه های ژنی می شود. اگر نرخ جابجایی ژنها کم باشد آنگاه نرخ کاهش در میزان ال دی با مقدار m یا نسبت مهاجرت در ارتباط خواهد بود. اگر نرخ مهاجرت ژنی بالا باشد در این صورت عامل تعیین کننده در کاهش میزان ال دی نسبت نو ترکیبی خواهد بود.

همخوانی:

همخوانی باعث تاخیر در کاهش عدم تعادل لینکاژ می شود زیرا این موضوع باعث کاهش فراوانی هتروزیگوت مضاعف در جمعیت می شود. این موضوع بخصوص در گونه های خود گشن به اثبات رسیده است.

انتخاب طبیعی:

انتخاب طبیعی باعث افزایش عدم تعادل لینکاژ بین ژن ها می شود و این موضوع از طریق کنار هم قرار دادن آلل ها در جایگاه های مختلف ایجاد می شود. واقعیت آن است که برخی از ژنها در صورت قرار گرفتن در کنار هم دارای اثر بالاتری نسبت به حالت طبیعی و پراکنش طبیعی در ژنوم هستند. بنابر این انتخاب طبیعی می تواند بعنوان عاملی که این ژنها را در کنار هم در یک ژنوم جمع می کند عمل نموده و باعث تولید سوپر ژنها شود. لینکاژ خیلی نزدیک باعث کاهش میزان نو ترکیبی بین ژنها در سوپر ژنها می شود و این موضوع باعث افزایش عدم تعادل لینکاژ در نتیجه انتخاب طبیعی خواهد شد.

نوترکیبی:

نوترکیبی باعث کاهش LD می شود. برای دو مکان ژنی که به صورت پیوسته در روی یک کروموزوم قرار دارند با در نظر گرفتن میزان نوترکیبی Θ مقدار عدم تعادل ژنتیکی بین آنها پس از هر نسل آمیزش تصادفی به مقدار $1-\Theta$ کاهش می یابد .

$$D_t = (1 - \Theta)^t D_0$$

در این رابطه Θ میزان نوترکیبی بین دو جایگاه ژنی و D_{MQ}^t عدم تعادل ژنتیکی بعد از t نسل است. طبق رابطه فوق عدم تعادل لینکاژ با افزایش میزان نوترکیبی در هر نسل نسبت به نسل های قبل کاهش می یابد و مقدار این کاهش با نرخ نوترکیبی مرتبط است. ضریب عدم تعادل لینکاژ D دارای دو خاصیت ناپسند است: اول آنکه این عدد بین -0.25 تا $+0.25$ تغییر می کند. حداکثر مقدار عدم تعادل گامتی هنگامی اتفاق می افتد که فراوانی گامت های $X11$ و $X22$ هر دو برابر با 0.5 باشد و یا اینکه فراوانی گامت های $X12$ و $X21$ هر دو برابر با 0.5 باشد. هنگامی که بین دو جایگاه که دارای عدم تعادل لینکاژ کامل هستند نوترکیبی بصورت آزاد رخ می دهد طی ۷ نسل عدم تعادل گامتی بطور کامل حذف می شود. بر این اساس با توجه به رابطه ذیل می توان تعداد نسل های مورد نیاز برای کاهش مقدار مشخصی در عدم تعادل گامتی را پیش بینی نمود.

$$t = [\ln (D_t/D_0)]/[\ln (1 - c)]$$

اگر در برخی مواقع LD بین دو جایگاه ژنی مشاهده و به ثبت برسد ممکن است این LD به مرور زمان بواسطه نوترکیبی از بین برود. البته میزان کاهش LD در صورتی که دو جایگاه ژنی بهم نزدیک باشند کمتر از زمانی است که دو جایگاه مورد بررسی دارای فاصله زیادی باشند. نرخ نوترکیبی در نواحی کروموزومی مختلف در گونه های مختلف متفاوت است. این موضوع یعنی تفاوت در نرخ نوترکیبی در بین افراد و نژادهای مختلف نیز وجود دارد. اگر برخی از این تغییرات بصورت ژنتیکی قابل کنترل باشد پس ما می توانیم انتظار داشته باشیم که نوترکیبی فرایندی است که در طول زمان افزایش می یابد. مطالعات متعدد نشان داده است که در یک دوره انتخاب در مگس دروزوفیل نرخ نوترکیبی تغییر کرده است. نتایج نشان داد که در صورت انتخاب برای سایر صفات مقدار نوترکیبی در دروزوفیل افزایش یافت. نرخ نوترکیبی در دو جنس نشان می دهد که معمولاً نرخ نوترکیبی در جنس ماده بالاتر از جنس نر است. در دروزوفیل هیچ نوترکیبی در نرها اتفاق نیفتاد. در انسان و در سلولهای غیر جنسی نرخ

نوترکیبی در ماده ها ۶۰ درصد بالاتر از نرها است. در جوامع بزرگ با آمیزش تصادفی میزان عدم تعادل ژنتیکی بین دو ژن پیوسته دارای فاصله فیزیکی نسبتاً زیاد تقریباً به نصف کاهش می یابد. برای مقادیر کمتر وقوع نوترکیبی بین ژنها، تعداد نسل لازم برای تغییر عدم تعادل به صفر بیشتر است.

چرا عدم تعادل لینکاژ و نوترکیبی برای ما مهم است؟

بخش قابل توجهی از واریانس ژنتیکی در هر نسل از طریق جور شدن مستقل کروموزوم ها ایجاد می شود. توجه داشته باشید که ژنوم انسان حاوی ۲۳ جفت کروموزوم است. بنابر این تعداد گامت های تولید شده در میوز با در نظر گرفتن جور شدن مستقل ژنها برابر است با:

$$2^{23} = 8,388,608.$$

حال اگر یک تبادل قطعه بین هر یک جفت کروموزوم همتا اتفاق بیفتد بنابر این در تقسیم میوز ۴ حالت مختلف گامت از این دو کروموزوم تولید می شوند. بنابر این تعداد گامت های متنوعی که با این فرض قابل تولید هستند برابر است با:

$$4^{23} = 7 \times 10^{13}$$

در نظر گرفتن همین تنوع در فرد دوم و سپس احتمال قرار گرفتن هر یک از این زیگوت ها از دو فرد در کنار هم باعث می شود تا تنوع افراد تولید شده به شرح ذیل باشد:

$$(4^{23})^2 = 5 \times 10^{27}$$

با این توضیحات شما متوجه می شوید که چرا نسبت به برادران و خواهرانتان متفاوت هستید. علاوه بر ایجاد تنوع قابل توجه در آمیزش تصادفی و در نتیجه انتخاب طبیعی، نوترکیبی باعث افزایش کارایی انتخاب طبیعی نیز می شود. دو اثر مفید که از نوترکیبی حاصل می شود حذف جهش های زیان آور و دیگری تثبیت جهش های مفید و سودمند است.

تفاوت مفهوم LD با مفهوم لینکاژ:

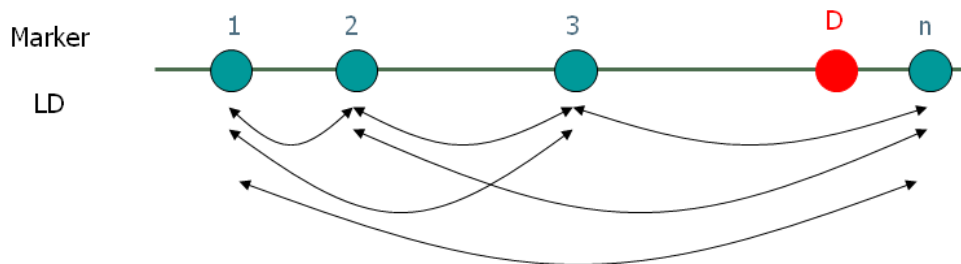
در ژنتیک جمعیت LD ارتباط غیر تصادفی آلله ها در دو یا چند جایگاه ژنی است که لزوماً نباید بر روی یک کروموزوم قرار داشته باشند. این اصطلاح را با عنوان عدم تعادل فاز گامتی و بصورت خلاصه عدم تعادل گامتی نیز می شناسند. به عبارتی LD ظهور برخی ترکیبات آلی

هاپلوتا‌پ) یا مارکرهای ژنتیک است که بیش تر یا کمتر از مقادیر مورد انتظار آنها (نسبت به حالت شکل گیری تصادفی هاپلوتا‌پ ها) در جمعیت ظهور یافته اند. مفهوم LD با مفهوم لینکاژ متفاوت است. لینکاژ به مفهوم ارتباط دو یا چند جایگاه بر روی کروموزوم است که بین آنها یک نوترکیبی محدود وجود دارد. مقدار LD به تفاوت بین فراوانی های آلی مشاهده شده و مورد انتظار بستگی دارد. جمعیتی که مقادیر مشاهده شده و مورد انتظار فراوانی آله‌ها در آن یکسان باشد در تعادل لینکاژ LE بسر می برد. مقدار عددی LD با توجه به لینکاژ ژنتیکی، انتخاب، میزان نوترکیبی، میزان جهش، گرایش تصادفی، آمیزش غیر تصادفی و ساختار جمعیت تغییر می کند. تاثیر نرخ نوترکیبی بر میزان LD در باکتری‌ها که دارای تکثیر غیر جنسی هستند قابل مشاهده است زیرا در این موجودات به دلیل عدم ایجاد نوترکیبی هیچگاه مقدار LD کاهش نمی یابد.

کاربردهای LD :

بطور کلی استفاده از LD در مطالعاتی که هدف آن بررسی ارتباط یک مارکر با مارکر دیگر و یا بررسی ارتباط یک مارکر با یک مکان ژنی و یا بررسی ارتباط یک مارکر با یک ژن بیماری‌زا است می توان از آماره های LD برای بررسی میزان ارتباط و پیوستگی مارکرها استفاده نمود. امروزه LD به عنوان یک فاکتور کلیدی در مطالعات آماری در مباحث ژنتیک مطرح است. مهمترین کاربردهای LD در مطالعات Genetic mapping , Genetic Association Study و Genomic Selection و QTL mapping است. مقدار LD یک پارامتر کلیدی در فرایند QTL mapping است. در QTL mapping از مارکرهایی که در LD نزدیک با جایگاه های ژن های کمی بسر می برند استفاده می شود تا جایگاه تقریبی یک ژن کمی در ژنوم تعیین شود. در مطالعات GS مارکرها نقش مهمی در محاسبه و توجیه واریانس ژنتیکی جمعیت مورد بررسی دارند. حداقل یک یا بیشتر مارکر SNP باید در LD با مکان QTL مورد بررسی باشد. اندازه نمونه با توجه به اندازه ژنوم و اندازه موثر جمعیت تعیین می شود که با LD ارتباط دارند. معیاری برای تعیین ارتباط مارکرهای ژنی و ژنهای کنترل مسبب بیماری است. مطالعات الگوی LD با استفاده از نشانگرها در جوامع زیستی باعث کشف و قربال ژنهای مسبب بیماری می شود. این بررسی

ها محل تقریبی جایگاه ژن D را تعیین می کند و اطلاعاتی را در مورد ساختار و پیچیدگی یک جایگاه ژنی ارائه می کند.



شانس نوترکیبی بین نواحی که LD بالایی دارند، کم است و بر عکس شانس وقوع نوترکیبی در نواحی که با هم LD پائین دارند، بیشتر است. با توجه به مقدار نوترکیبی اتفاق افتاده بین دو جایگاه ژنی، فاصله ژنها بر اساس سانتی مورگان مشخص شده و تقدم و تاخر آنها در ژنوم تعیین می شود.

چه نوع مارکرهای برای اندازه گیری LD مناسب اند؟

از آنجا که LD نسبت به لینکاژ در فواصل ژنتیکی کوچک تری بسط یافته است لذا برای کشف و تشخیص LD باید از تراکم بالاتری از مارکرها استفاده شود. دو نوع مارکر مورد استفاده در مطالعات ژنتیک شامل مارکرهای میکروستلایت و SNP ها هستند. در گاو حدود ۲۱۴۱ مارکر میکروستلایت وجود دارد که با توجه به اینکه ژنوم گاو ۳۰۰۰ سانتی مورگان طول دارد لذا در هر ۱.۴ سانتی مورگان یک مارکر میکروستلایت وجود دارد و لذا در گاو مارکرهای میکروستلایت برای بررسی LD می تواند کارایی داشته باشد.

در انسان میکروستلایت مارکر مناسبی برای بررسی LD نیست چون از تراکم کافی برخوردار نیست. اما SNP هم در انسان و هم در دام های اهلی و سایر حیوانات بعنوان یک نشانگر متراکم (۱) عدد SNP در هر ۱۰۰۰ باز (از قابلیت بالایی برای مطالعات LD برخوردار است. عیب SNP ها آن است که دو آللی بوده و واریانس را بخوبی مارکرهای میکروستلایت نشان نمی دهند و لذا در ازای هر میکروستلایت به ۵ عدد SNP برای بررسیهای LD نیاز است.

مقایسه LD در جمعیت های انسانی و دامی:

اگر بپذیریم که میزان LD با اندازه موثر جمعیت ارتباط مستقیم دارد، باید گفت که طبق مشاهدات اندازه موثر جمعیت در جوامع انسانی بسیار بزرگتر از جمعیت های دامی است. اندازه موثر جمعیت در انسان حدود ۱۰۰۰۰ نفر (Kruglyak, 1999) و در گاو بعنوان مثال حدود ۱۰۰ راس است (Riquet et al. 1999).

بر اساس اندازه موثر جمعیت مقدار LD باید در جوامع انسانی بسیار پایین تر از جمعیت های دامی باشد. معمولاً با توجه به نوع جمعیت مورد مطالعه بسط LD در یک جمعیت انسانی در فواصل ۵ کیلوباز (تقریباً 0.005 سانتی مورگان) مشاهده شده است. در گاو گوسفند مقدار LD قابل توجهی در فواصل ۵ تا ۱۰ سانتی مورگانی بسط یافته است (Franir et al. 2000, McRae et al. 2002, Hayes et al. 2003).

آماره های اندازه گیری و بیان میزان LD :

۱- آماره D (Lewontin 1960): با در نظر گرفتن دو مکان ژنی A, B

LD خصوصیت توزیع هاپلوتایپ ها در مورد دو جایگاه ژنی است و ارتباط بین دو جایگاه ژنی بر روی دو کروموزوم را در جمعیت نشان می دهد. فرض صفر در باره آن است که تعادل لینکاژ در جمعیت بین دو جایگاه ژنی در دو کروموزوم همولوگ برقرار بوده و فراوانی هاپلوتایپ ها در جمعیت با دقت از روی فراوانی های آللی قابل محاسبه است. با فرض دو مکان ژنی A و B هر یک با دو آلل مقدار آماره D با توجه به فراوانی های هاپلوتایپی و ژنی به روش ذیل محاسبه می شود:

	Locus B		Totals	
	B	b		
Locus A	A	p_{AB}	p_{Ab}	p_A
	a	p_{aB}	p_{ab}	p_a
Totals		p_B	p_b	1.0

$$D = P(AB) - P(A).P(B)$$

در این رابطه $P(AB)$ عبارت است از فراوانی هاپلوتایپ AB و P_A و P_B عبارت است از فراوانی آلل های A و B . نکته قابل توجه آن است که اگر مقدار D بر اساس فراوانی سایر هاپلوتایپ ها نیز محاسبه شود، یک عدد واحد حاصل می شود.

روش دیگر محاسبه D براساس فراوانی انواع شکل های متفاوت گامتی آنها می باشد. مطابق قانون هاردی واینبرگ اگر جامعه در حال تعادل ژنتیکی باشد فراوانی انواع متفاوت گامت ها باید یکسان باشد. به عنوان مثال، فرض شود یک نشانگر با دو آلل M_1 و M_2 به صورت پیوسته با یک مکان ژنی کمی با دو آلل Q_1 و Q_2 باشد. برای این دو ژن در جامعه چهار نوع گامت متفاوت M_1Q_1 و M_2Q_1 و M_1Q_2 و M_2Q_2 می تواند وجود داشته باشد که فراوانی آنها به ترتیب u ، s ، t و r در نظر گرفته شود. مقدار لینکاژ ژنتیکی نامتعادل در این حالت برابر است با:

$$D_{MQ} = ru - st$$

معایب آماره D

آماره D دارای دو مشکل عمده است. اول آنکه مقدار D به فراوانی های ژنی وابسته است و این موضوع باعث می شود تا مقایسه LD در مکان های ژنی مختلف با استفاده از این آماره دشوار باشد. و دوم آنکه تفسیر این آماره مشکل است چون با توجه به مقدار D برآورد شده نمی توان گفت که آیا مقدار LD در حداکثر مقدار ممکن است یا خیر؟

۲- آماره D استاندارد: D'

با توجه به معایب ذکر شده برای آماره D بنابر این Lewontin در ۱۹۶۴ مقدار D استاندارد را بر اساس حداکثر مقادیر ممکنه با فرمول ذیل ارائه کرد:

$$D' = D / D_{\max}$$

که در این فرمول $D_{\max}$ حداکثر مقدار D در رابطه با فراوانی آللی مورد نظر است.

$$D_{\max} = \begin{cases} \min(p_1q_1, p_2q_2) & \text{when } D < 0 \\ \min(p_1q_2, p_2q_1) & \text{when } D > 0 \end{cases}$$

D' بین صفر و یک تغییر می کند و امکان بیان مقدار عدم تعادل لینکاژ را بر اساس حداکثر مقدار عدم تعادل لینکاژ فراهم می کند. اگر مقدار D' برابر ۱ باشد نشان می دهد که مقدار LD حداکثر مقدار ممکن است و اگر برابر با صفر باشد نشان می دهد که LD بین دو جایگاه ژنی وجود ندارد و دو جایگاه نسبت به هم مستقل هستند و نوترکیبی بصورت کامل و متعادل اتفاق افتاده است.

معایب معیار D' :

۱- اگر فراوانی یک آلل در جمعیت بسیار کم باشد (آلل نادر) ممکن است معیار D' بیش از

حد واقع برآورد شود و در نتیجه مقدار LD نیز بیش از حد برآورد می شود.

۲- اگر اندازه نمونه کوچک باشد مقدار D' بیش از حد برآورد می شود.

علت موارد فوق آن است که در چنین شرایطی ممکن است مارکر از قدرت کافی برای

تشخیص هاپلوتایپ چهارم برخوردار نباشد.

۳- آماره r^2 (Hill و Robertson در سال ۱۹۶۸)

آقایان Hill و Robertson در سال ۱۹۶۸ مقادیر ذیل را برای اندازه گیری عدم تعادل لینکاژ بیان کردند.

$$r^2 = D^2 / (p_1 \cdot p_2 \cdot q_1 \cdot q_2)$$

مقدار r^2 ممکن است بصورت r که همان میزان همبستگی بین دو جایگاه است:

$$r = \frac{D}{\sqrt{p_1 p_2 q_1 q_2}}$$

مقدار r^2 بین ۰ و ۱ تغییر می کند. شبیه آنچه که در مورد مقدار D گفته شد مقدار حداکثر r^2 به

فراوانی های آللی بستگی دارد. اگر مقدار r^2 برابر با ۱ باشد نشانه آن است که دو جایگاه در LD کامل

هستند و میزان همبستگی بین دو جایگاه نیز ۱۰۰ درصد است. اگر مقدار r^2 برابر با ۰ باشد نشانه آن است که دو جایگاه کاملاً از هم مستقل اند و بین آنها LD وجود ندارد. ما می توانیم آلل ها را بعنوان متغیرهای تصادفی کمی در نظر بگیریم که در محدوده ۰ تا ۱ تغییر می کنند. پس R^2 ضریب همبستگی یک جفت الل است. در حقیقت مقدار واقعی ضریب عدم تعادل لینکاژ از توزیع احتمال فرایندهای تکاملی استخراج شده است.

اگر هدف ما تعیین یک پلی مورفیسم باشد بهتر است از معیار r^2 بجای سایر معیارها استفاده کنیم زیرا این آماره مقدار تغییر در عملکرد را در مطالعات همبستگی نشان می دهد (وقتی که آلل A جایگزین آلل B می شود). این آماره اعداد بین ۰ و ۱ را می پذیرد. r^2 برابر ۱ خواهد بود اگر و تنها اگر فراوانی هر دو هاپلوتایپ مشاهده شده برابر صفر باشد. مقدار R^2 برابر صفر نشان می دهد که دو آلل مورد بررسی در تعادل کامل بسر می برند.

معایب آماره r^2 :

این آماره برای محاسبات جفتی مناسب نیست. زیرا تصور نحوه محاسبه ضریب r^2 بین چندین جایگاه مختلف دشوار است. این آماره برای نشانگرهای چند آللی عمومیت ندارد. و محاسبه آن بر مبنای آن است که توزیع هاپلوتایپ های جمعیت شناخته شده است. مقدار χ^2 که در قبل توصیف شد در واقع فراوانی برآورد شده از هاپلوتایپ ها است و مقدار واقعی آنها نیست. اگر اندازه نمونه محدود باشد و برآورد فراوانی های هاپوتایپ اشتباه باشد مقدار r^2 ممکن است گمراه کننده باشد.

تفاوت و شباهت آماره D' و r^2 در بیان LD

شباهت:

هر دو آماره ای برای بیان و اندازه گیری مقدار LD بین دو جایگاه هستند.

تفاوت:

مقدار r^2 همبستگی بین دو جایگاه را نشان می دهد ولی اطلاعاتی راجع به وقوع یا عدم وقوع نوترکیبی ارائه نمی کند. اما مقدار D' رابطه معکوس با نوترکیبی دارد و $D'=1$ به مفهوم آن است که

بین دو جایگاه هیچ نوترکیبی رخ نداده است. ولی اطلاعاتی راجع به میزان همبستگی دو جایگاه ارائه نمی کند.

تست χ^2 :

هنگامی که ما یک فایل داده هاپلوتایپ را برای محاسبه LD آنالیز می کنیم ما می خواهیم فرض صفری که دو جایگاه دارای تعادل لینکاژ هستند را تست کنیم. به یاد داشته باشید که آماره χ^2 برای یک نمونه هاپلوتایپ یا کروموزوم می تواند بصورت ذیل محاسبه شود:

$$\chi_s^2 = \sum_{i=1}^k \frac{(m_i - Np_i)^2}{Np_i}$$

در رابطه بالا m_i فراوانی مشاهده شده برای هاپلوتایپ ها است. P_i احتمال وقوع هاپلوتایپ i از توزیع مورد فرض است. N اندازه نمونه یا تعداد کروموزوم است. K تعداد هاپلوتایپ موجود در توزیع است (مثلا در مورد نشانگرهای دو آللی مقدار K برابر با ۴ است).

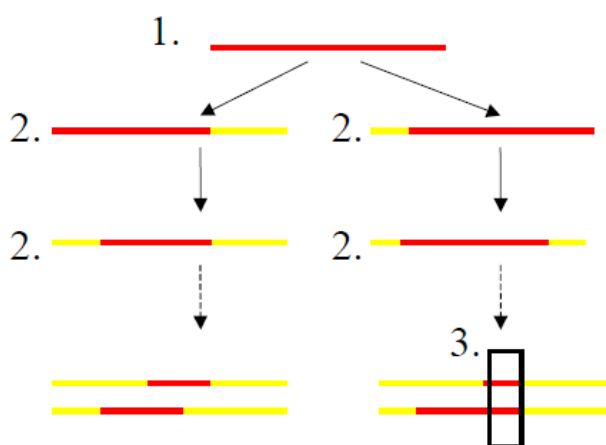
۵-آماره CSH:

این آماره در سال ۲۰۰۳ توسط (Hayes et al. 2003) ارائه شد. تعریف های کلاسیک از مقدار LD بصورت عملی ممکن است در مطالعات نقشه یابی QTL مورد استفاده قرار نگیرد. برای مثال آماره r^2 در یک زمان فقط دو جایگاه ژنی را مد نظر قرار می دهد در حالیکه ما می خواهیم بسط LD را در تمام قطعات کروموزومی تعیین کنیم که نیاز به بررسی چندین مارکر بطور همزمان دارد. لذا یک روش جایگزین برای محاسبه LD بین چندین مارکر آماره CSH است که مخفف واژه Chromosome Segment Homozygosity (تشابه قطعات کروموزوم) است.

توجه داشته باشید که اجداد یک جمعیت در نسل های گذشته دارای فرزندان در نسل کنونی هستند. اگر کروموزوم اجداد طی نسل های مختلف دچار نوترکیبی شده باشند ممکن است هنوز بخش کوچکی از کروموزوم های اجداد در نسل های کنونی به همان صورت قبلی وجود داشته باشند که

وجه تشابه افراد نسل حاضر هستند. به این نواحی کروموزومی اصطلاحاً identical by descent

(IBD) می گویند. Identical By Descent (IBD)



بر این اساس تعریف CSH احتمال مشابه و مشترک المنشا بودن (از یک جد منشأ بگیرند) دو قطعه کروموزوم در دو فرد از جمعیت است. که هیچ گونه نو ترکیبی در ساختار این دو قطعه ایجاد نشده است.

نحوه محاسبه CSH:

مقدار CSH برای بخشی از یک کروموزوم یا تمام طول یک کروموزوم قابل ارائه و محاسبه است. مقدار CSH بصورت مستقیم از روی مارکرها قابل برآورد نیست. اما بصورت غیر مستقیم از روی نشانگرهای هاپلوتایپ مربوط به یک قطعه کروموزوم قابل برآورد است. فرض کنید یک قطعه کروموزوم با مارکر A و قطعه دیگر با مارکر B مشخص شده و این دو در کنار هم هاپلوتایپ AB را تشکیل داده اند. احتمال هموزیگوت بودن این هاپلوتایپ (HH) از رابطه ذیل محاسبه می شود.

$$HH = CSH + \frac{(Hom_A - CSH)(Hom_B - CSH)}{1 - CSH}$$

بر این اساس مقدار CSH می تواند با معلوم بودن میزان هموزیگوسیت هاپلوتایپ (HH) و میزان هموزیگوسیت انفرادی هر یک از مارکرها A و B که از روی فایل اطلاعات حاصل می شود، محاسبه شود.

محاسبه میزان عدم تعادل لینکاژ با حل یک تمرین ساده:

در جدول زیر مقادیر مشاهده شده از هاپلوتایپ های مربوط به دو جایگاه ژنی A , B ارائه شده است:

	B1	B2	
A1	۱	۰	۱
A2	۴۹	۵۰	۹۹
	۵۰	۵۰	۱۰۰

مقادیر D ، D' و r^2 را بین این دو جایگاه محاسبه کنید؟

الف- محاسبه مقدار D :

$$D = p(A_1B_1) - p(A_1)p(B_1)$$

$$2N=100$$

$$P(A_1) = (n_{11} + n_{12}) / 2N = (1 + 0) / 100 = 0.01$$

$$P(B_1) = (n_{11} + n_{21}) / 2N = (1 + 49) / 100 = 0.5$$

$$P(A_2) = (n_{21} + n_{22}) / 2N = (49 + 50) / 100 = 0.99$$

$$P(B_2) = (n_{12} + n_{22}) / 2N = (0 + 50) / 100 = 0.5$$

$$p(A_1B_1) = 1 / 100 = 0.01$$

$$D = 0.01 - (0.01)(0.5) = 0.005$$

مقدار D مثبت و مخالف صفر است. این عدد نشان می دهد که بین دو جایگاه A و B عدم تعادل

لینکاژ وجود دارد. اما این عدد هیچ توضیحی در رابطه با حداقل یا حداکثر بودن مقدار LD بیان نمی

کند.

ب- محاسبه D' :

$$D' = D / D_{\max}$$

$$D_{\max} = \begin{cases} \min(p_1q_1, p_2q_2) & \text{when } D < 0 \\ \min(p_1q_2, p_2q_1) & \text{when } D > 0 \end{cases}$$

برای محاسبه D' با توجه به اینکه مقدار D مثبت و بزرگتر از صفر است، لذا مقدار حداقل $p_{A1}p_{B2}$ ، $p_{A2}p_{B1}$ بعنوان $D_{\max}$ در محاسبه D' مورد استفاده قرار می گیرد:

$$P(A_1) \cdot P(B_2) = 0.01 \times 0.5 = 0.005 \quad \text{min}$$

$$p(A_2) \cdot P(B_1) = 0.99 \times 0.5 = 0.495$$

$$D' = (0.01 - (0.01)(0.5)) / (0.005) = 1$$

مقدار D' برابر با یک است و این عدد نشان می دهد که LD بین دو جایگاه A و B در حداکثر مقدار ممکن است و هیچ نوترکیبی بین دو جایگاه ژنی وجود ندارد.

ج- محاسبه r^2 :

$$r^2 = D^2 / (p_{A1} p_{A2} p_{B1} p_{B2})$$

$$= (0.005)^2 / (0.01 * 0.99 * 0.5 * 0.5) = 0.002475$$

$$r = 0.0497$$

مقدار r بین دو جایگاه A , B مثبت و مخالف صفر و برابر ۰/۰۴۹ است. این عدد نشان می دهد که اولاً بین دو جایگاه LD وجود دارد و ثانياً همبستگی بین این دو جایگاه پائین است.

برخی از نرم افزارهای برآورد LD :

GOLD - "Graphical Overview of LD"(Abecasis and Cookson)

Arlequin (Schneider, Roessli, Excoffier)

Haplo (Kidd et al.)

MIDAS: software for analysis and visualization of inter allelic disequilibrium between multi allelic markers

LDcompare

R (package genetics)

BMAPBUILDER(Building chromosome-wide LD maps)

BOOLD(for calculations under different linkage disequilibrium (LD) distribution models)

CHROMSCAN(for association mapping of disease genes)

DMLE(mapping of the position of a disease mutation)

EMLD(estimation of haplotype frequencies and LD calculation)

آشنایی با برنامه LdCompare

LdCompare برنامه ای برای تعیین میزان LD در مطالعات ژنومیکس است. این برنامه می تواند میزان LD را برای نشانگرهای چندگانه و منفرد محاسبه نموده و همچنین همپوشانی ژنتیکی آنها را تعیین نماید. مقیاس اطلاعات بدست آمده از مطالعات ژنو میکس روزبه روز در حال افزایش است (نزدیک ۹ میلیون SNP در بانک اطلاعات dbSNP تا کنون به ثبت رسیده است).

<http://www.ncbi.nlm.nih.gov/SNP>; (Karchin et al., 2005; Wang et al., 2005)]

اخیرا در پروژه بین المللی HAPMAP تعداد ۴ میلیون SNP در ۴ گروه نژادی مختلف ژنوتایپ گردید. این پیشرفت ها امکان مطالعه گسترده و جامع الاطراف ارتباط ژنها در سطح ژنوم را فراهم آورده است. امروزه به لحاظ اقتصادی و فنی تعیین تمام SNP ها در یک نمونه امکان پذیر نیست. خوشبختانه ارتباط بین نشانگرهای نزدیک بهم (که معیار تعیین کننده آن LD است) به ما اجازه می دهد تا زیر مجموعه ای از تمام SNP ها را بررسی نمائیم و از روی آنها ژنوتایپ SNP های مجاور را استنتاج نمائیم. بعلاوه پیشرفت های اخیر در بیوتکنولوژی امکان تعیین ژنوتایپ ۵۰۰ هزار یا بیشتر SNP را در یک آزمایش فراهم آورده است. بصورت موازی در روش های آماری سعی بر آن است تا با استفاده از مدل های مناسب داده های حاصل از این مطالعات با هدف بررسی ارتباطات ژن ها در سطح ژنوم مورد تجزیه و تحلیل قرار بگیرد. یک جنبه مهم در این ارتباط برآورد میزان LD موجود بین SNP ها و همپوشانی ژنتیکی آنها است. بصورت خلاصه همپوشانی قطعات بخشی از SNP ها در مخزن ژنی مورد مطالعه است که همبستگی آنها با حداقل یکی از SNP های تعیین ژنوتایپ شده بیشتر از حد آستانه موجود باشد. همپوشانی نشانگرهای چندگانه و منفرد (با استفاده از یک نشانگر یا مجموعه ای از نشانگرها برای نشانه گیری SNP ها در اطلاعات هدف) بصورت گسترده در برآوردهای آماری مورد استفاده قرار می گیرند. هرچند این محاسبات در تعداد زیاد SNP باعث ایجاد پیچیدگی خواهد شد. برای مثال در فاز دوم پروژه HAPMAP (مطالعه ۴ میلیون SNP) که یک مورد مطالعه ساده برای تعیین همپوشانی مارکرهای منفرد بود تعیین تمام همبستگی های جفتی

مارکرها در محدوده ۱۰۰ کیلوبازی نیازمند حدود 10^8 محاسبه برای r^2 بود. و این یعنی پیچیدگی محاسبات. در محاسبات پیچیده تر برای تعیین همپوشانی بین سه مارکر به محاسبه 10^{13} مقدار برای r^2 نیاز مندیم. بعلاوه تعداد محاسبات در صورت افزای فضای مورد مطالعه بصورت اکسپوننشال افزایش خواهد یافت. با توجه به دانسته های ما برنامه LDCompare اولین نرم افزاری است که می تواند همپوشانی سه مارکری را در کل ژنوم برآورد نماید.

خلاصه ای از خصوصیات برنامه:

برنامه دارای محیط کار واحدی برای مارکهای منفرد و چندگانه است. و می تواند بصورت اتوماتیک با توجه به فایل داده ها نوع روش را تشخیص دهد. اطلاعات وارد شده می تواند بصورت ژنوتایپ دیپلوئید و یا بصورت اطلاعات هاپلوتایپ فازی به برنامه معرفی شود. R^2 جفتی و مطالعات همپوشانی با استفاده از مارکهای منفرد در هر دو نوع داده ورودی قابل محاسبه است اما محاسبه r^2 نشانگرهای چندگانه و همپوشانی آنها نیازمند اطلاعات هاپلوتایپ فازی است (زیرا تهیه ساختارهای ۳ و ۴ تایی از SNP هاپلوتایپ ها بسیار زمان بر و غیر دقیق است).

فرمت استاندارد فایل داده های مربوط به لینکاژ مکان های ژنی که برای داده های دیپلوئید مورد استفاده قرار می گیرند در آدرس ذیل قابل دسترسی است.

<http://www.broad.mit.edu/mpg/haploview/files.php>

که از این فایل ما برای تولید آلوگوریتم ای ام به منظور تولید هاپلوتایپ های دو مارکری استفاده می کنیم.

خروجی برنامه مربوط به r^2 مارکهای منفرد و چندگانه: محققان می توانند از این اطلاعات برای محاسبات توزیع تجمعی همپوشان و انتخاب SNP های نزولی استفاده کنند.

برنامه در زبان C++ نوشته شده است و امکان مشارکت استفاده کنندگان در بازونویسی و تکمیل برنامه وجود دارد.

کدینگ برنامه بگونه ای است که سریع عمل کرده و حجم حافظه کمی نیاز دارد. این برنامه در محیط ویندوز و لینوکس قابل استفاده است. این برنامه در رایانه ای با سی پی یو ۲/۸ گیگا هرتز intel و با ۱ ram گیگا هرتز قابل استفاده است. این برنامه می تواند تمام ژنوم را برای r^2 جفتی و همپوشانی نشانگرهای منفرد با انحراف معیار ۱۰۰ هزار جفت باز را در ۲ ساعت محاسبه کند. محاسبه همپوشانی

دو مارکری در این برنامه حدود ۱۲ ساعت طول می کشد. بررسی همپوشانی سه مارکری به زیر گروه هایی برای کروموزوم های مختلف تقسیم می شود و این کار برای همه کروموزوم ها بصورت همزمان و موازی انجام می شود. برای بزرگترین کروموزوم ها این فرایند ۲۴ ساعت بطول می انجامد.

محاسبه LD بین سه مارکر با استفاده از نرم افزار R

فرض کنید ژنوتیپ افراد برای سه جایگاه مارکری مختلف در جدول ذیل ارائه شده است:

Compute pair-wise linkage disequilibrium between
genetic markers

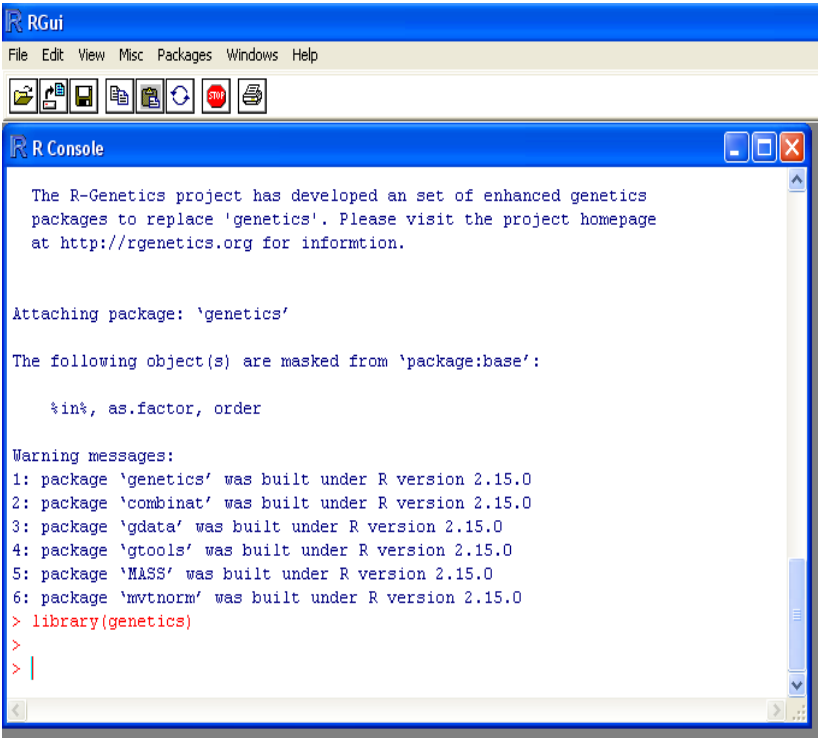
Genotype1	Genotype2	Genotype3
"T/A"	"C/A"	"T/A"
"T/A"	"C/A"	"T/T"
"T/T"	"C/C"	"T/A"
"T/A"	"C/A"	"T/T"
"T/T"	"C/C"	"T/A"
"T/A"	"C/A"	"T/T"
"T/A"	"C/A"	"T/T"
"T/A"	"C/A"	"T/A"
"T/A"	"C/A"	"A/A"
"T/T"	"C/C"	"T/T"
"T/A"	"C/A"	"T/A"
"T/T"	"A/A"	"T/A"
"T/A"	"C/A"	"T/T"
"T/A"	"A/A"	"T/A"
"T/A"	"C/A"	"T/A"
"T/T"	"C/C"	"T/A"
"T/A"	"C/A"	
"T/A"	"C/A"	
"T/A"	"C/A"	
"T/T"	"A/A"	

نرم افزار R را می توان از سایت اختصاصی آن و از CRAN R به راحتی و به صورت رایگان دانلود کرد. این نرم افزار دارای بسته های نرم افزاری اختصاصی است که هر یک در زمینه خاصی از علوم نظیر آمار، ریاضی، زیست شناسی، زراعت، ژنتیک و اصلاح نژاد، علوم اجتماعی و ... کاربرد دارند. این بسته ها را بسادگی می توان از سایت اختصاصی و CRAN R دانلود کرد. بسته نرم افزاری genetics یکی از بسته های اختصاصی نوشته شده در زمینه ژنتیک است که قابلیت های متعددی داشته که یکی از این قابلیت ها برآورد مقادیر LD است.

در کنار این بسته نرم افزاری دستورالعمل استفاده از آن به همراه فایل های اجرایی قابل دانلود است که استفاده از این نرم افزار را ساده تر می کند. پس از دانلود نرم افزار و نصب آن در صفحه اجرایی R با استفاده از دستورات مراحل ذیل را در پیش می گیریم:

۱- فراخوانی پکیج genetics:

با تایپ دستور `library(genetics)` بسته نرم افزاری genetics را که قبلا نصب شده است فراخوانی می کنیم.



```
RGui
File Edit View Misc Packages Windows Help

R Console

The R-Genetics project has developed an set of enhanced genetics
packages to replace 'genetics'. Please visit the project homepage
at http://rgenetics.org for informtion.

Attaching package: 'genetics'

The following object(s) are masked from 'package:base':

    %in%, as.factor, order

Warning messages:
1: package 'genetics' was built under R version 2.15.0
2: package 'combinat' was built under R version 2.15.0
3: package 'gdata' was built under R version 2.15.0
4: package 'gtools' was built under R version 2.15.0
5: package 'MASS' was built under R version 2.15.0
6: package 'mvtnorm' was built under R version 2.15.0
> library(genetics)
>
> |
```

۲- معرفی ژنوتیپ مارکر ها به نرم افزار

در مرحله بعد با استفاده از دستور `genotype` ژنوتیپ افراد را در رابطه با سه مارکر مورد بررسی با نام های `g1,g2,g3` به نرم افزار معرفی می کنیم:

```

RGui
File Edit View Misc Packages Windows Help

R Console

%in%, as.factor, order

Warning messages:
1: package 'genetics' was built under R version 2.15.0
2: package 'combinat' was built under R version 2.15.0
3: package 'gdata' was built under R version 2.15.0
4: package 'gtools' was built under R version 2.15.0
5: package 'MASS' was built under R version 2.15.0
6: package 'mvtnorm' was built under R version 2.15.0
> library(genetics)
>
>
> g1 <- genotype( c("T/A", NA, "T/T", NA, "T/A", NA, "T/T", "T/A",
+ "T/T", "T/T", "T/A", "A/A", "T/T", "T/A", "T/A", "T/T",
+ NA, "T/A", "T/A", NA) )
>
> g2 <- genotype( c("C/A", "C/A", "C/C", "C/A", "C/C", "C/A", "C/A", "C/A",
+ "C/A", "C/C", "C/A", "A/A", "C/A", "A/A", "C/A", "C/C",
+ "C/A", "C/A", "C/A", "A/A") )
>
> g3 <- genotype( c("T/A", "T/A", "T/T", "T/A", "T/T", "T/A", "T/A", "T/A",
+ "T/A", "T/T", "T/A", "T/T", "T/A", "T/A", "T/A", "T/T",
+ "T/A", "T/A", "T/A", "T/T") )
>
> |

```

۳- محاسبه LD بین مارکر g1 و g2:

در مرحله بعد با استفاده از دستور LD(g1,g2) مقدار LD بین دو مارکر g1 , g2 محاسبه

می شود و خروجی ذیل تولید می شود:

```

RGui
File Edit View Misc Packages Windows Help

R Console

>
>
> g1 <- genotype( c("T/A", NA, "T/T", NA, "T/A", NA, "T/T", "T/A",
+ "T/T", "T/T", "T/A", "A/A", "T/T", "T/A", "T/A", "T/T",
+ NA, "T/A", "T/A", NA) )
>
> g2 <- genotype( c("C/A", "C/A", "C/C", "C/A", "C/C", "C/A", "C/A", "C/A",
+ "C/A", "C/C", "C/A", "A/A", "C/A", "A/A", "C/A", "C/C",
+ "C/A", "C/A", "C/A", "A/A") )
>
> g3 <- genotype( c("T/A", "T/A", "T/T", "T/A", "T/T", "T/A", "T/A", "T/A",
+ "T/A", "T/T", "T/A", "T/T", "T/A", "T/A", "T/A", "T/T",
+ "T/A", "T/A", "T/A", "T/T") )
>
>
> LD(g1,g2)

Pairwise LD
-----
              D          D'       Corr
Estimates: 0.1419402 0.8110866 0.6029553

              X^2      P-value   N
LD Test: 10.90665 0.0009581958 15

> |

```

در خروجی بالا مقدار آماره های D و D' و r (تحت عنوان corr) ارائه شده است. همچنین

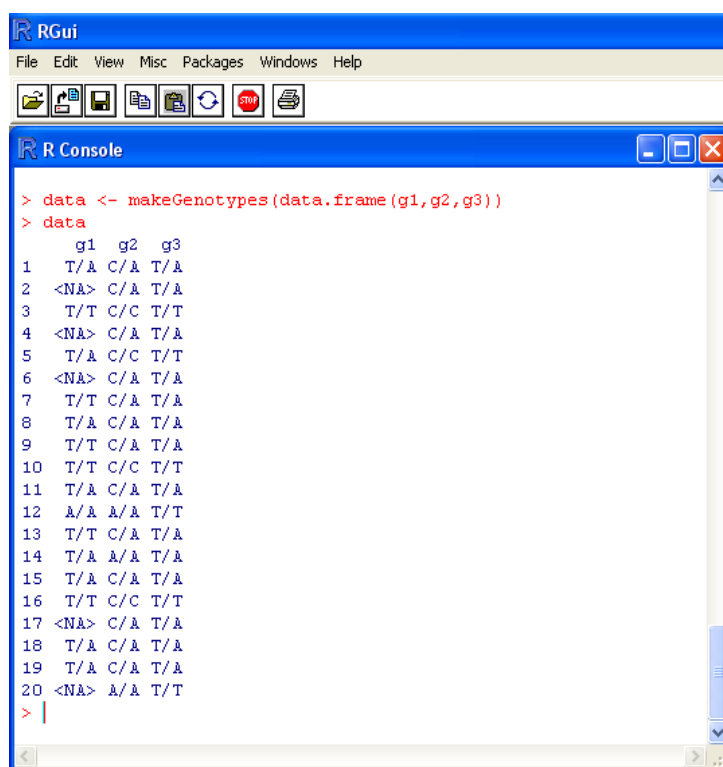
با استفاده از آزمون k اسکور فرض صفر عدم وجود LD بین دو مارکر آزمون شده است که با توجه

به مقدار p-value مشخص است که فرض صفر رد شده و این نشان می دهد که مقدار LD بین دو جایگاه معنی دار است.

مقدار D برابر با ۰/۱۴ است و این عدد نشان می دهد که بین دو جایگاه LD وجود دارد. مقدار D' نشان می دهد که مقدار LD نسبتاً بالا است و معادل ۸۱ درصد مقدار حداکثر LD را شامل می شود. همچنین مقدار آماره r^2 نشان می دهد که همبستگی ژنوتیپ ها در دو جایگاه مارکری نسبتاً بالا و معادل ۶۰ درصد است.

محاسبه LD دو به دو بین سه مارکر

تهیه ژنوتیپ از سه مارکر مختلف: برای تهیه ژنوتیپ از سه مارکر مطرح شده در مثال فوق در نرم افزار R از دستور makeGenotypes استفاده کرده و سه ژنوتیپ را به این ترتیب به نرم افزار معرفی می کنیم:



```

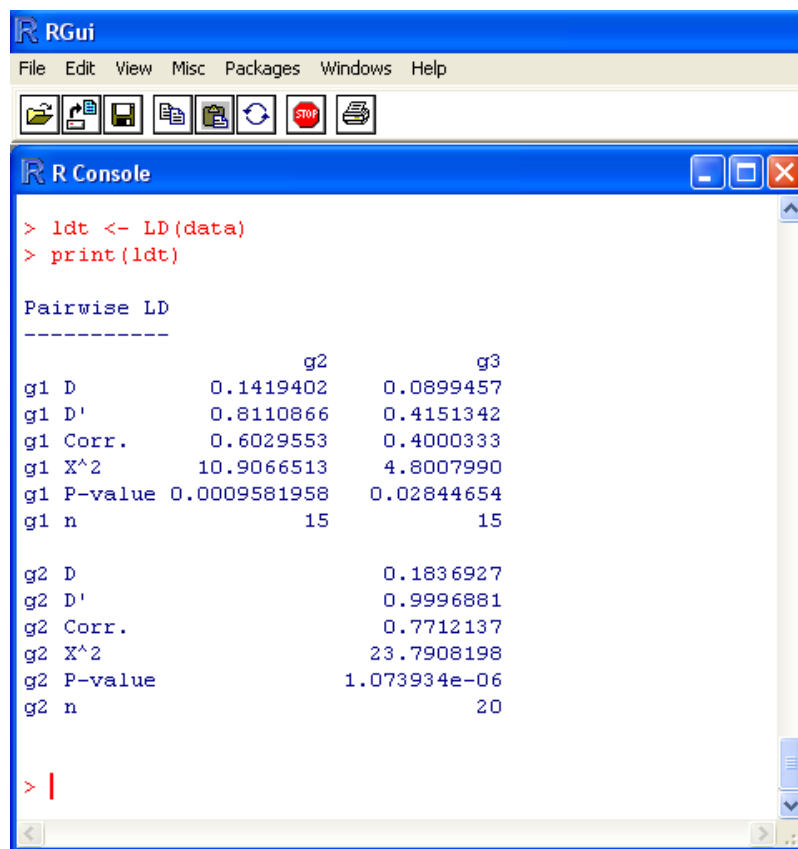
R RGui
File Edit View Misc Packages Windows Help

R Console
> data <- makeGenotypes(data.frame(g1,g2,g3))
> data
  g1  g2  g3
1  T/A C/A T/A
2 <NA> C/A T/A
3  T/T C/C T/T
4 <NA> C/A T/A
5  T/A C/C T/T
6 <NA> C/A T/A
7  T/T C/A T/A
8  T/A C/A T/A
9  T/T C/A T/A
10 T/T C/C T/T
11 T/A C/A T/A
12 A/A A/A T/T
13 T/T C/A T/A
14 T/A A/A T/A
15 T/A C/A T/A
16 T/T C/C T/T
17 <NA> C/A T/A
18 T/A C/A T/A
19 T/A C/A T/A
20 <NA> A/A T/T
  
```

محاسبه LD بین سه مارکر مختلف

سپس با استفاده از دستور LD(data) مقدار LD بین سه مارکر را محاسبه کرده و نتیجه را در

فایلی به نام ldt ذخیره می کنیم. دستور print(ldt)



```
> ldt <- LD(data)
> print(ldt)
```

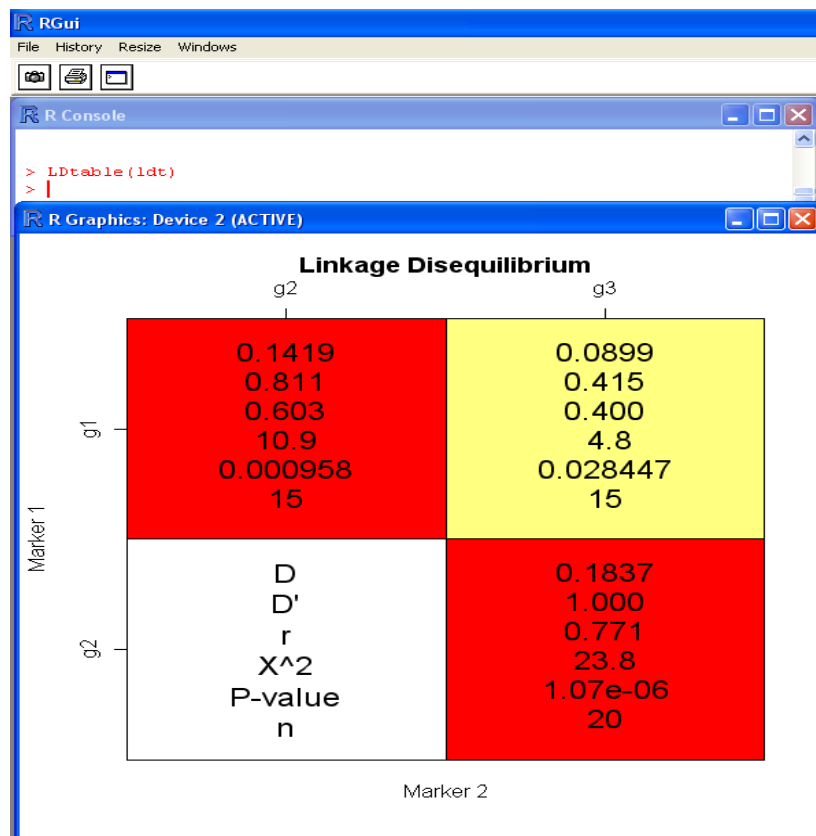
Pairwise LD

	g2	g3
g1 D	0.1419402	0.0899457
g1 D'	0.8110866	0.4151342
g1 Corr.	0.6029553	0.4000333
g1 X^2	10.9066513	4.8007990
g1 P-value	0.0009581958	0.02844654
g1 n	15	15
g2 D		0.1836927
g2 D'		0.9996881
g2 Corr.		0.7712137
g2 X^2		23.7908198
g2 P-value		1.073934e-06
g2 n		20

در خروجی بالا مقادیر $D, D', r, p\text{-value}, K^2$ برای ترکیب های دو تایی از مارکرها ارائه شده است. برای مثال در خروجی بالا مشخص است که میزان LD بین دو مارکر g_2, g_3 برابر با ۹۹/۹۶ مقدار Ld حداکثر بوده و از LD بین دو مارکر g_1, g_2 و همچنین بین دو مارکر g_1, g_3 بیشتر است. مقادیر r و $p\text{-value}$ نیز همین واقعیت را نشان می دهند.

رسم نمودار رنگی LD بین سه مارکر

برای تهیه نمودارهای رنگی از مقادیر LD بین سه مارکر مورد بررسی می توان از دستور LDtable در بسته genetics استفاده کرد. با تایپ این دستور جدول رنگی مربوط به میزان LD بین مارکرها چاپ می شود:



در جدول فوق رنگ تیره تر (قرمز) نسبت به رنگ روشن تر (زرد) نشانه وجود LD بیشتر بین دو مارکر است. در هر مربع جدول نیز مقادیر آماره های مختلف LD و مقدار p-value درج شده است.

شبیه سازی مارکرهای بیشتر

با استفاده از تابع های موجود در نرم افزار R می توان برنامه ای را نوشت که تعداد بیشتری مارکر را شبیه سازی نماید. مثلاً در دستور ذیل ۲۰ مارکر شبیه سازی شده و مقدار LD دو به دوی آنها بصورت جدول رنگی ارائه شده است.

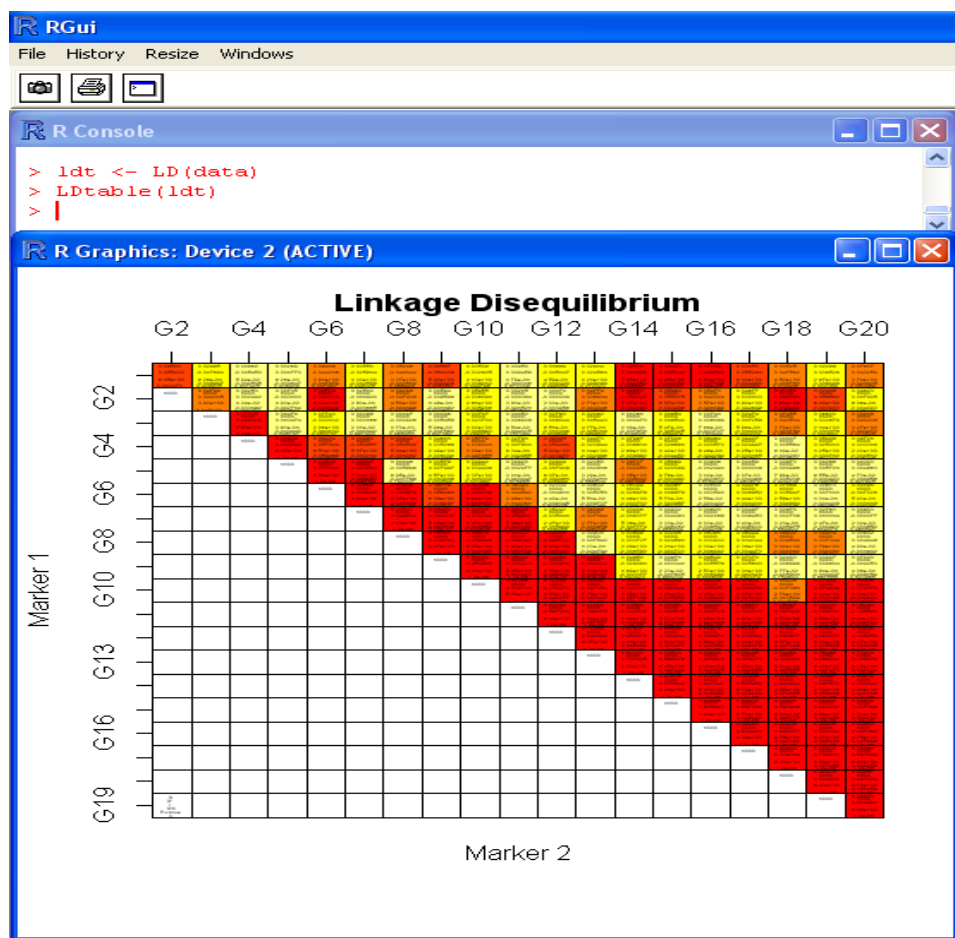
با تایپ دستور LD(data) و سپس تایپ دستور LDtable می توان جدول رنگی LD بین این ۲۰ مارکر را نمایش داد. در هر خانه جدول ارقام مربوط به آماره های بیان کننده LD تایپ شده است و خانه هایی با رنگ تیره تر نشان دهنده مقدار LD بیشتر بین دو مارکر است.

```

RGui
File Edit View Misc Packages Windows Help

R Console
> # more markers makes prettier plots!
> data <- list()
> nobs <- 1000
> ngene <- 20
> s <- seq(0,1,length=ngene)
> a1 <- a2 <- matrix("", nrow=nobs, ncol=ngene)
> for(i in 1:length(s) )
+ {
+   rallele <- function(p) sample( c("A","T"), 1, p=c(p, 1-p))
+   if(i==1)
+   {
+     a1[,i] <- sample( c("A","T"), 1000, p=c(0.5,0.5), replace=TRUE)
+     a2[,i] <- sample( c("A","T"), 1000, p=c(0.5,0.5), replace=TRUE)
+   }
+   else
+   {
+     p1 <- pmax( pmin( 0.25 + s[i] * as.numeric(a1[,i-1]=="A"), 1 ), 0 )
+     p2 <- pmax( pmin( 0.25 + s[i] * as.numeric(a2[,i-1]=="A"), 1 ), 0 )
+     a1[,i] <- sapply(p1, rallele )
+     a2[,i] <- sapply(p2, rallele )
+   }
+   data[[paste("G",i,sep="")]] <- genotype(a1[,i],a2[,i])
+ }
> data <- data.frame(data)
> data <- makeGenotypes(data)
> |

```



- A Compilation of Some Available Software for Linkage disequilibrium analysis. <http://www.mrc-epid.cam.ac.uk/Personal/jinghua.zhao/software.htm>.
- Branca, Timothy D. Paape, Peng Zhou, Roman Briskine, Andrew D. Farmer, Joann Mudged, Arvind K. Bharti, Jimmy E. Woodward, Gregory D. May, Laurent Gentzbittele, Cécile Bene, Roxanne Denny, Michael J. Sadowsky, Joëlle Ronfort, Thomas Bataillon, Nevin D. Young, and Peter Tiffin. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. PNAS | October 18, 2011 | vol. 108 | no. 42.
- Gaunt, T.R., S. Rodriguez, C. Zapata and I.N.M. Day. 2006. MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics* 2006, 7:227.
- Georges, M. 2010. Linkage disequilibrium: a primer. *Unit of Animal Genomics University of Liège*.
- Goldstein, D. B., M. E. Weale. 2001. Population genomics: Linkage disequilibrium holds the key. *Current Biology* Vol 11 No 14:576-579.
- Hardy–Weinberg principle. From Wikipedia, the free encyclopedia. This page was last modified on 31 January 2012 at 00:22.
- Hayes, B. 2003. Linkage disequilibrium mapping (LD) mapping and Bioinformatics Approaches to identify QTL. Department of Primary Industries Victoria, Australia.
- Hayes, B. 2007. QTL Mapping, MAS, and Genomic Selection. A short-course organized by Animal Breeding & Genetics Department of Animal Science, Iowa State University. June 4-8, 2007.
- Hayes, B., Peter M. Visscher, Helen C. McPartlan, and Mike E. Goddard. 2003. Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Research* 13:635.
- Hao, K., X. Di and S. Cawley. 2006. LdCompare: rapid computation of single- and multiple-marker r^2 and genetic coverage. *BIOINFORMATICS APPLICATIONS NOTE*. Vol. 23 no. 2 2007, pages 252–254.

Huttley, G. A., Michael W. Smith, Mary Carrington and Stephen J. O'Brien.1999. A Scan for Linkage Disequilibrium Across the Human Genome. Genetics 152: 1711–1722 .

Istrail, S.2010. Introduction to the r^2 Measure for Linkage Disequilibrium.

Li, Na and Matthew Stephens.2003. Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. Genetics 165: 2213–2233.

Linkage disequilibrium. From Wikipedia, the free encyclopedia. This page was last modified on 16 February 2012 at 14:50.

Warnes, G. R. 2010.Pairwise linkage disequilibrium between genetic markers. LD {genetics}. R Documentation. gregory_r_warnes@groton.pfizer.com.