Definitions

- **.fsa File** - A raw data file generated during sample electrophoresis as part of a run; only viewable through GMID software (or equivalent).
- **Administrative documentation** - Materials associated with Case Records which do not include technical records but may include scanned copies of additional Request for Physical Examination of Evidence Forms, internal chain of custody documents, Forensic Scientist statement of qualifications (CV), notes and communication logs of case-related conversations, subpoenas and records of discovery, Sexual Assault Evidence Collection Kit papers, Subject Evidence Collection Kit papers, and other pertinent information which relates to the Case Record, but does not necessarily support the conclusions drawn.
- **Allele -** An alternative form of a gene; allele designation is used to represent a specific size fragment of DNA for a specific locus in STR analysis.
- **Allelic Dropout -** Failure to detect an allele within a sample or failure to amplify an allele during PCR.
- **Amelogenin** – Gender-determining locus.
- **Analytical Threshold (AT)-** The minimum height (RFU) requirement at and above which detected peaks can be reliably distinguished from background noise; peaks above this threshold are generally not considered noise and are either artifacts or true alleles. The threshold for this Laboratory is internally derived by empirical data.
- **Artifact** – Non-allelic byproducts of PCR technology (e.g., stutter, etc.), anomalies which occur during capillary electrophoresis (e.g., pull-up, spike, etc.), or byproducts of primer synthesis (e.g., dye blob, etc.).
- **Bank** - A specific set of samples whose position indicates to the QIAgility where to place those samples into the reaction plate.
- **Bin** – An expected location for a particular allele within a locus; a binset is a collection of expected locations for alleles at loci amplified as a set.
- **Body swab** – Intimate swab collected from a person, not from an internal body cavity. Normally submitted as a part of sexual assault evidence.
- **Carrier RNA** – Component of the DNA Investigator Kit that is present to enhance the binding of DNA to the silica surface of the magnetic particles present in the reagent cartridge. This enables more efficient isolation of low amounts of DNA from samples.
- **Combined technical and administrative review** – An evaluation of reports, notes, data, and supporting documentation to ensure that there is an appropriate and sufficient basis for the scientific conclusions as well as consistency with Laboratory policies and editorial correctness.
- **Commercial Reagent**: A commercially produced laboratory reagent designed to conduct a specific forensic test. All commercial reagents shall have an expiration date either established by the manufacturer or, if none is provided, the Forensic Biology Section shall establish the expiration date. Commercial reagents in the Forensic Biology Section: ProK (both stock supply and aliquots), ProK from DNA Investigator Kit, DTT (both stock supply and aliquots), Hi-Di formamide (both stock supply and aliquots), 20 % SDS, 10x buffer, 1x buffer, nuclease-free dH2O, WEN sizing standard, spectral/matrix kits for 3500 (or equivalent) and 7500 (or equivalent), Carrier RNA (cRNA) from the DNA Investigator Kit.
- **Composite Profile:** A DNA profile generated by combining typing results from different loci obtained from multiple injections of the same amplified sample and/or multiple amplifications of

Form Approved for use by:

DNA Technical Leader

the same DNA extract. When separate extracts from different locations on a given evidentiary item are combined prior to amplification, the resultant DNA profile is not considered a composite profile.

- **Consensus Allele** – An allele that conforms to an incremental repeat pattern at an STR locus.
- **Core Loci:** The 20 loci defined by the FBI and required for inclusion within CODIS. The 20 core loci are CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, and D22S1045.
- **Critical Instrument:** Determined by empirical studies or routine practice to require reliability testing on established samples before use on evidentiary samples. Critical instruments shall have performance checks prior be being placed into service and routinely thereafter. Critical Instruments in Forensic Biology: Qiagen EZ1 Advanced, QIAcubes, QIAgility, AB 7500s, Proflex thermal cyclers, AB 3500/3500XL genetic analyzers.
- **Critical Reagent:** Determined by empirical studies or routine practice to require reliability testing on established samples before use on evidentiary samples. All critical reagents shall have an expiration date as established by the manufacturer or the Forensic Biology Section. Critical reagents in the Forensic Biology Section: STR-Tris-EDTA (STR-TE), STR-Stain Extraction Buffer (STR-SEB), commercially supplied kits and their components (PowerPlex® Fusion 6C and PowerPlex® Y23, Quantifiler® Trio, DNA Investigator).
- **CT --** The cycle number when fluorescence in the assay crosses the threshold of detection and the PCR reaction is in the exponential phase; lower CT values indicate samples with higher concentrations, and higher CT values indicate samples with lower concentrations.
- **Derived (deduced) contributor:** A distinct contributor profile generated by using PHR and mixture interpretation and/or the application of a reference standard to the original mixture profile
- **Differential lysis**: An extraction process used for the analysis of mixed stains. It includes what is referred to as sperm and non-sperm lysis. The non-sperm lysis is contained in the aqueous portion remaining after a gentle lysis treatment of the stain. A more rigorous treatment follows for the pellet material (generally sperm); this is referred to as the sperm lysis.
- **Distinguishable Mixture:** A mixture in which relative peak height ratios allow for the determination of a major contributor(s). Separation of contributors (into major and minor components) is based on quantitative peak height information (see Peak Height Ratio).
- **Distinguishable portion of the case file number** - At a minimum, the case number shall reflect the year and the five numbers unless one of the five numbers is a place holding zero.
- **Dithiothreitol (DTT)**: A chemical present to reduce the disulfide bonds that maintain the integrity of the sperm head. Sperm heads do not readily lyse in the absence of DTT.
- **DNA Investigator Kit**: Supplies provided by Qiagen for the extraction of forensic casework. This kit includes, the reagent cartridge, Proteinase K, carrier RNA, tubes and tips.
- **DNA Profile:** The combination of genotypes obtained from DNA analysis testing of multiple loci.
- **Electronic Record** - A data file that has information recorded in a form that only a computer can process.
- **Electropherogram (egram)** - The computer generated electronic/visual result from an analysis performed by electrophoresis.

Form Approved for use by:

DNA Technical Leader

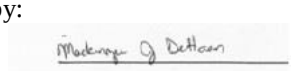
- **Ethylenediaminetetraacetic acid (EDTA)**: A component of the reactions used in the lysis process, which inhibits nuclease activity.
- **Evidence -** An object submitted to the State Crime Laboratory for analysis. An item of evidence is equivalent to a test item as described in ISO 17025.
- **Examination documentation** – Records of tests conducted, standards and controls used, diagrams, printouts, photographs, spectra, chromatograms, hand-written notes and other material used by the Forensic Scientist to reach a conclusion.
- **Exclusion:** A conclusion reached after comparing the DNA profile of a known sample to the DNA profile of an evidentiary item and the individual in question is not a potential contributor.
- **Form** - A document with a fixed arrangement of captioned spaces designed for entering and extracting information. Forms become a record once completed.
- **Full Profile:** A DNA profile that exhibits genotypic information at each locus tested and there is no evidence of allelic dropout, degradation, or preferential amplification.
- **Genotype:** Characterization of the alleles present at a genetic locus; the combination of genotypes obtained for multiple loci is referred as a DNA profile
- **Haplotype:** Combination of alleles from several loci on the Y chromosome that are inherited together.
    - **Heterozygote:** Two different alleles at a particular gene locus on homologous chromosomes.
    - **Homozygote:** Same alleles at a particular gene locus on homologous chromosomes.
- **Inclusion:** A conclusion reached after comparing the DNA profile of a known sample to the DNA profile of an evidentiary item and the DNA profile of the individual in question is a potential contributor. Due to the inherited nature of haplotypes, paternal relatives of an individual would also be included to the evidentiary sample.
- **Inconclusive profile/component:** A DNA typing result which stems from an insufficient quantity/quality of DNA, (e.g., degraded DNA, preferential amplification, stochastic effects, and/or number of contributors). This type of profile provides insufficient information to support an inclusion or exclusion and shall not be used for comparison purposes.
- **Indistinguishable from Stutter:** A peak located within a stutter location that is of similar RFU value to other unambiguous minor alleles at other loci. These peaks are not conclusive since they meet both the expectations of stutter and of being an allele.
- **Indistinguishable Mixture:** A mixture in which the relative peak height ratios do not allow for the determination of a major contributor(s).
- **Inhibition** – The total or partial suppression of the PCR process that would result in a partial or no DNA profile being obtained.
- **Internal Positive Control (IPC)** – A synthetic sequence of DNA not found in nature that is present in all Quantifiler Trio reactions. It is used as a means to distinguish between negative sample results and reactions affected by the presence of PCR inhibitors, assay setup, or chemistry/instrument failure.
- **Intimate Sample:** A biological sample from an evidence item that is obtained directly from an individual's body; it is not unexpected to detect that individual's allele(s) in the DNA typing results (e.g. items collected as a part of a sexual assault evidence collection kit and fingernail scrapings).
- **Known**: Biological material whose identity is established. These types of samples are used in casework for comparison to evidence.

Form Approved for use by:

DNA Technical Leader

- **Limited Access:** A location to which not all Forensic Scientists, Supervisors, or Managers have access.
- **LIZ Sizing Standard** - A commercially produced set of DNA fragments of known size, used to determine the size of DNA fragments in an electrophoresed sample.
- **Locus (plural = Loci)** - The chromosomal location or location of a gene or DNA marker.
- **Major Contributor:** An unambiguous single primary source of DNA within a mixture as determined by the application of the PHR and other mass parameters.
- **Match:** DNA profiles are considered to match if their patterns are the same after taking into consideration the properties of the substrate tested and limitations of the specific techniques used. (See inclusion definition.)
- **Microvariant** - An allele that varies by less than the consensus repeat unit and is not defined by a ladder allele. Microvariants are observed in-between the ladder alleles for a specific locus.
    - **Minimum allele frequency:** A frequency which shall be used for any allele which is observed 5 or less times in the population frequency database (to include variant and off-ladder alleles). The formula for this frequency is 5/2N (2N = # of alleles in the population database at that locus).
- **Minor contributor**: A secondary source of DNA within a mixture as determined by the application of PHR, number of contributors, and interpretation of the data generated
- **Mixture:** A DNA typing result originating from more than one individual.
- **Multiple Major Contributors:** The presence of more than one predominant contributor to a mixture profile.
- **Noise** - Signal detected by a data collection instrument.
- **Non Template Control (NTC)** – A negative control used in the quantitation assay; STR-TE is used.
- **Off Ladder Allele** - Alleles that size outside the allele categories (bins) represented in the ladder.
- **Off-Scale Data** - The result of excess DNA present in an electrophoresed sample, typically visualized by excessive artifacts as a result of peak heights consistently greater than 12,000 RFUs. The determination of off-scale data must take into account dose of the allele being evaluated (homozygote vs heterozygote). The heterozygote peaks should be summed to determine contribution for that locus.
    - **Orifice swab**- intimate swab collected from an internal body cavity (e.g. vaginal, rectal, and oral). Normally included as a part of a Sexual Assault Evidence Collection Kit.
    - **Panel** – A collection of markers specific to an amplification kit (i.e., PowerPlex® Fusion 6C).
- **Partial DNA Profile:** A DNA profile that does not produce DNA typing results for all loci tested due to DNA degradation, inhibition, or low quantity DNA template.
- **Peak Height Ratio (PHR):** The relative ratio of two alleles at a given locus, as determined by dividing the peak height of an allele with a lower relative fluorescence unit (RFU) value by the peak height of an allele with a higher RFU value and then multiplying the value by 100 to express the PHR as a percentage.
- **Popstats –** The computer program used to generate Random Match Probability (RMP) and CPE/CPI statistics.

Form Approved for use by:

DNA Technical Leader 

- **Portable Document Format (PDF)** – A file format that preserves most attributes of a source document no matter which application, platform, and hardware type was originally used to create it.
    - **PowerPlex® Fusion 6C** - A commercially produced amplification kit which contains the following loci: CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, Penta D, Penta E and SE33 as well as Amelogenin, DYS391, DYS570, DYS576 for gender determination.
- **PowerPlex® Y23** – A commercially produced amplification kit which contains the following loci: DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 and Y-GATA-H4.
- **Predominant DNA Profile/Major Contributor:** An unambiguous single primary source of DNA within a mixture as determined by the application of the PHR.
    - **Product Rule:** A calculation that can be used if a locus is inherited independently of the other loci. The calculation can be made by multiplying each individual genotype frequency together.
- **Project** – A set of data saved as a single entity in GMID-X.
- **Proteinase K (ProK)**: A proteolytic enzyme that reduces proteins to their constituent amino acids. In particular, ProK removes the histone groups that keep the DNA tightly bound within the cell. The enzymatic activity of ProK lasts for approximately two hours and eventually self-digests.
- **Pull-up:** A signal from an allele labeled with one dye-set which may show up as a peak or Off-Ladder Allele in another dye-set.
- **QCO**: Refers to the DNA Quality Control Officer and/or his/her designee.
- **Questioned Sample:** Biological sample recovered from a crime scene or collected from persons or objects associated with a crime.
- **R^2 value** – This value measures the closeness of fit between the standard curve regression line and the individual CT data points of quantification standard reaction. A value of 1.00 indicates a perfect fit between the DNA data points and the curve.
- **Random Match Probability (RMP):** the probability of randomly selecting an unrelated individual from the population who could be a potential contributor to an evidentiary profile.
- **Reagent Cartridge**: Reagents for the purification of nucleic acids from a single sample on the EZ1 Advanced XL BioRobot are contained in a single reagent cartridge. Each well of the cartridge contains a particular reagent, such as magnetic particles, buffers, and water.
- **Records** - Materials created or received by the Laboratory that are preserved as documentation of the activities of the Laboratory or for the value of the information. Records include, but are not limited to, reports, correspondence, telephone logs, quality records and technical records.
- **Reference Sample:** Biological material for which the identity of the donor is established and used for comparison purposes; also referred to as a known standard. These include victim, suspect (subject), elimination and/or witness standards.
- **Relative Fluorescence Units (RFUs)** – A unit of measurement in electrophoresis when fluorescence detection is used; determines peak height.

Form Approved for use by:

DNA Technical Leader

  - **Restricted:** Referring to a statistical approach conditioned on the number of contributors and with consideration of quantitative peak height information and inference of contributor mixture ratios; used to limit the genotypic combinations of possible contributors.

- **Run/Injection** – Each set of samples that is injected and separated electrophoretically on a Capillary Electrophoresis Unit (24 for a 3500xL Genetic Analyzer or 8 for a 3500).
- **Secondary Reference Sample**: Due to the inherited nature of Y-STR, in cases where a reference sample from the listed individual is unavailable (e.g. missing person cases), a standards from a known paternal relative may be used as an alternate standard.
- **Shoulder and Tail:** A Shoulder and Tail is an elongated or raised area to the immediate left and right of a main peak but is not separated from the main peak.
- **Single Source Profile:** A combination of genotypes obtained from STR DNA testing that could originate only from a single individual (taking into account paternal relatives due to the inherited nature of Y-STRs). A sample may be considered to consist of a single contributor when no more than two alleles are observed at each locus. All loci are to be evaluated in making this decision. If three alleles are observed at one locus, then there may not be a mixture; the individual contributor may have a triallelic pattern at that locus.
  - **Sizing Standard** - A commercially produced set of DNA fragments of known size, used to determine the size of DNA fragments in an electrophoresed sample.
- **Slope** – This value indicates the efficiency of the PCR reaction in the assay. A value of -3.3 indicates 100 % efficiency.
- **Sodium dodecyl sulfate (SDS)**: A chemical whose presence serves to rupture the cell nuclear membrane to expose the nucleic acids. It also assists in the denaturation of the nuclear proteins that are attached to the DNA.
- **Spike/Electrical Spike:** An artifact believed to be caused by an increase in the current within a capillary that causes a sharp increase in signal. This artifact lacks the defined morphology of a peak.
- **Split Peaks** - A split peak is where one allele is represented by two peaks. Lack of full nucleotide A addition may be observed when the amount of input DNA is greater than the recommended protocol. In this case, more time is needed for Taq Polymerase to add the A nucleotide to all molecules. Amplification of too much input DNA also results in off-scale/overblown data (saturation of signal) and may be manifested as split peaks.
- **Stochastic Effects:** The observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity template samples.
- **Stochastic Threshold (ST):** The value above which it is reasonable to assume allelic drop-out has not occurred within a single source sample. The threshold for this Laboratory is internally derived through the use of empirical data. Used to assist in evaluating intuitiveness of genotype combinations.
- **Stutter:** An artifact of PCR amplification resulting from strand slippage during amplification. (typically one repeat unit less (e.g. N-4) or one repeat unit more (e.g. N+4) than the corresponding main allele peak, additional repeats are also seen based off of repeat length of corresponding STR).

Form Approved for use by:

DNA Technical Leader

- **Technical records -** Accumulations of data and information which result from performing tests as specified in technical procedures. Technical records include, but are not limited to, forms, worksheets, photographs, and test reports.
- **Triallelic Pattern:** Three peaks observed at a single locus and not the result of a mixture. These peaks may or may not be of equal intensity.
- **Unincorporated Dye:** Unincorporated dye (i.e., dye-blobs) may be observed in an electropherogram and are distinct morphologically from a labeled DNA fragment. A dye-blob does not exhibit the typical sharp, distinct peak that is produced by actual alleles and is observed as a wider, thicker peak and may be lacking the sharply defined slope to the apex of a peak.
- **Unknown**: Biological material whose identity has not been established. These types of samples are used in casework for comparison to any available known samples. Also called "questioned" item.
    - **Unrestricted:** Referring to a statistical approach without consideration of quantitative peak height information and inference of contributor mixture ratios; for CPE/CPI this may or may not be conditioned on the number of contributors.
- **Work product -** The material that is generated as a function of the analysis and includes DNA extracts, amplified product, amplification tubes, 96 well plates used in analysis.
- **Worktable** – The location where the user loads samples and the components of the assay kit.
- **Y-intercept** – The Ct value of a theoretical 1 ng/µl sample; a shift of 1 CT can result in either a half or a doubling of the estimation of a sample's quantity.

STRmix™ Definitions:

STRmix^{TM}: a fully continuous probabilistic genotyping forensic software program which combines biological modeling with mathematical processes in order to (1) interpret and attempt to deconvolute DNA profiles in the presence or absence of conditioned samples, and (2) provide statistical weight in the form of a likelihood ratio (LR) after the comparison of reference samples to evidence samples.

- **Accepts or Moves –** an iteration that is accepted and 'moved' towards the next step in the MCMC process. This is defaulted in STRmix^{TM} to 400,000 accepts (8 chains – 50,000 per chain)

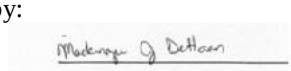
- **Acceptance rate -** total number of accepts during the MCMC process, divided by total iterations.

- **Allele Variance (c^2) -** Parameter that describes how variable allele peak heights are within the run

- **Average (log) likelihood -** this value is the average log10(likelihood) for the entire post burn-in MCMC.

- **Burn-in**—During the MCMC process, the first accepts are discarded. This is done to allow the chains time to reach a desired or 'good space'. Default value is 10,000 per chain.

Form Approved for use by:

DNA Technical Leader 

- **Conditioned sample/reference-** A sample where a known reference sample is used during the deconvolution of the contributors. Application of a conditioned reference will better inform the genotype combinations selected during the MCMC process. A reference may be conditioned on samples where the presence of an individual's DNA on an item is expected (e.g., intimate sample, samples collected from person's clothing or often used item.)

- **Deconvolution** - process performed using a Markov Chain Monte Carlo process which creates possible genotype combination(s). Each combination is assigned a weight which reflects how well it explains the evidence profile.

- **Degradation –** estimated for each contributor, the decreasing trend of peak heights with increasing molecular weight – a higher number indicates a steeper slope of degradation.

- **Degradation starting at (rfu/bp) -** the lowest RFU peak at the smallest molecular weight locus in the sample.

- **Detection threshold -** lab-specific analytical threshold (AT) determined during internal validation studies.

- **DNA Amounts (RFU) -** The best estimate of the "amount" (proportion) of DNA, expressed in RFU, determined by the mean of all post burn-in accepts of the calculated template (t) for each contributor

- **Drop-in cap –** maximum summed heights of drop-in allele(s) (in RFU) permitted per locus. STRmix^{TM} will not model an allele as drop-in if it is above the drop-in cap.

- **Drop-in frequency –** laboratory observed rate of drop-in observed during internal validation

- **Drop-in parameters –** parameters such as analytical threshold, peak height, observed drop-in rate, and probability of drop-in that are used to model drop-in in STRmix^{TM}; described as a gamma distribution ($\alpha 3$, $\beta 3$).

- **Effective Sample Size thinning –** the number of values STRmix^{TM} uses in the ESS calculation if there were 2 million iterations and the ESS thinning setting is 1 million then the ESS calculation will be performed by thinning out every second value. This assists with run time. If the number of iterations is less than the ESS thinning value then STRmix^{TM} uses all of them.

- **Effective Sample Size (ESS) -** the number of independent samples that have been taken from the posterior distribution of the MCMC likelihood.

- **Factor of N! –** When this value is set to "yes", the DNA profile of a comparison sample (i.e. victim or suspect) is compared to the mixture as a whole. When this value is set to "no", the DNA profile of a comparison sample is compared to each component of the mixture.

Form Approved for use by:

DNA Technical Leader

- **Gelman-Rubin convergence diagnostic -** this value informs the user whether the MCMC chains have converged. This is calculated by comparing the within-chain and between-chain variances of the MCMC chains.

- **Highest Posterior Density (HPD) –** defines the interval most likely to contain the true value; used when calculating a likelihood ratio

  - **HPD iterations –** the number of iterations used within the Highest Posterior Density calculation to create the probability interval. The default setting is 1000.
  - **HPD Significance value –** the percentile used within the HPD calculation for the probability interval. The default setting is 99.0.
  - **HPD Sides –** The number of sides used within the HPD calculation for the probability interval (1 or 2).

- **99.0% 1-sided lower HPD -** The combined LR calculated by STRmix™ is referred to as a point estimate. Because the true answer is not known, a confidence interval is then applied around the point estimate known as a 1-sided 99.0% highest HPD credible interval. This interval accounts for the uncertainty associated with the point estimate LR. This interval, commonly applied in Bayesian statistical calculations, gives a range (i.e. with 99% confidence) of where the true allele probabilities actually lie. The lower end of the HPD interval is reported from STRmix™ to be the most conservative to the person of interest.

- **Iteration** – a proposed genotype combination that is either accepted or rejected during the MCMC process. These proposed genotype combinations are not the final result.

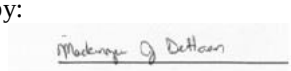
- **Likelihood Ratio (LR)-** calculated by comparing the probabilities of two hypotheses H1 and H2 (Hp and Hd, in STRmix^{TM}). STRmix^{TM} incorporates the assigned weights and sub-population models (Balding and Nichols, 1994, also known as NRC II recommendation 4.2) to calculate the LR.

- **LR total -** combined LR for all loci.

- **Locus amplification variance (LSAE variance, σ^2) -** A variable that penalizes loci which have locus amplification efficiencies that differ from the mean of the other loci values. The locus amplification variance parameter was determined through the internal validation using Model Maker and is expected to be affected by laboratory specific variables. It is used to correct for variation in amplification efficiencies in order to pull locus expected peak heights back toward s the whole profile expected heights.

- **Locus efficiencies (also referred to as Locus Specific Amplification Efficiencies, or LSAE) -** how well each locus in the sample amplified in comparison to other loci within the sample; this is modeled as one of the variable parameters within the MCMC process

Form Approved for use by:

DNA Technical Leader 

- **Mass parameters**- variable parameters used to generate expected peak heights during a deconvolution, collectively referred to as **DART**:
    - **D**egradation rate for each contributor to the DNA profile
    - **A**mplification efficiencies for each locus within the profile
    - **R**eplicate amplification strength for each PCR replicate
    - **T**emplate for each contributor (measured in relative fluorescence units (RFU)

- **MCMC** - Markov Chain Monte Carlo is a mathematical method that uses a random re-sampling process in order to give a best explanation for an observed set of data
    - **Markov chain:** A process that steps from one position to another. In STRmix™, the positions define certain combinations of genotypes and parameters that are being tried as explanations for the profile. At each iteration, the algorithm may either step or stay put. The chain is the path of steps and stay puts.
    - **Markov property**: the property of the MCMC analysis where the values in each step are independent of the values in the previous step.
    - **Monte Carlo:** a random re-sampling method used to address complex problems. It is used to systematically search the entire range of possibilities that are being considered to ensure that all likely combinations are considered. In STRmix™, this range of possibilities consists of series of proposed genotypes for each given locus.

- **MCMC Accepts**- The number of MCMC acceptances required (including burn-in) before the MCMC finishes. The default value is 50,000 per chain for post-burn in, 10,000 for burn-in (400,000 total).

- **MCMC Chains**- a Markov chain is a process that steps from one position to another. The positions define certain combinations of genotypes and parameters that are being tried as explanations for the profile. At each decision the algorithm may either step to a new position (accept) or stay (reject the new position) at the same position. The chain is the path of steps and stays.

- **MCMC Uncertainty –** When set to "yes", STRmix™ considers genotype set weights, allele frequencies and theta as distributions and re-samples from these distributions during the LR HPD calculation. This acts as an additional layer of conservatism.

- **Maximum degradation –** the maximum allowable degradation for any one contributor during burn-in.

- **Maximum stutter –** the maximum allowable back stutter proportion permitted, i.e. 0.30 = 30%. Setting stutter max = 0 turns this parameter off. The maximum stutter parameter should be determined by the laboratory through internal validation studies and is expected to be affected by laboratory specific variables.

- **Metropolis-Hastings algorithm:** The Metropolis-Hastings algorithm is an equation used to

Form Approved for use by:

DNA Technical Leader

determine whether or not to accept or reject a new profile combination.

- **Minimum allowed Variance from the mode –** The minimum allowable value the allele and stutter variance constants can take in relation to the mode of their prior distributions. For example, if the mode is 4.2 and the minimum allowed variance from the mode setting is 0.5, then the smallest value the variance constant can take is 4.2 x 0.5 = 2.1.

- **Mixture Proportions (%) -** approximate percentage of each contributor to the sample.

- **Model Maker –** A function within the software that provides an estimation of the STRmix™ parameters for an STR amplification kit using empirical data as a part of the laboratory's validation studies. The parameters within Model Maker must be determined before casework samples can be analyzed in STRmix™.

- **Per Locus Likelihood Ratios -** For each locus, the probability for the evidence given H1 (H_p) and H2 (H_d) are individually listed, as well as the ratio of the two probabilities (LR).

- **Primary diagnostics-** diagnostics that can be intuitively approximated by an experienced analyst such as weights, LRs and mix ratio/proportion.

- **Q allele –** signifies a genotype possibility that includes a dropped allele which is not observed in the evidence profile.

- **Random Walk Standard Deviation-** sets the step size distributions for the random Gaussian walks. During the MCMC, the next iteration will be close but not too close to the previous iteration.
    - When RWSD is small – STRmix™ takes smaller steps which leads to more accepts and is quicker.
    - When RWSD is large – STRmix™ takes larger steps which leads to less accepts and is slower, but will more likely step across valleys.

- **Relationship LRs-** LR which considers the evidence being explained in H2 (H_d) by a relative of the comparison sample in H1 (H_p).

- **Saturation –** All data above this level is only considered qualitatively. For example, if Saturation = 12000 RFU and the evidence input file contains some peaks with heights >12000 RFU, STRmix™ will convert these peak heights to 12000 RFU and run accordingly. The value for saturation is determined during internal validation studies and is expected to be specific to the model of electrophoresis instrument used. Samples that are saturated should be evaluated with caution.

- **Secondary diagnostics-** diagnostics that need to be evaluated based on STRmix™ output rather

Form Approved for use by:

DNA Technical Leader

than intuitive knowledge such as allele variance, stutter variance, average log likelihood, and Gelman-Rubin diagnostic

- **Seed value -** starting number used within the random number generator. Setting the seed to the same number will return the same results for a sample from run to run and should only be used to assist with validation and performance checks.

- **Stratified LR –** when multiple populations are selected to calculate an LR, STRmix^{TM} will calculate LRs for each population individually and then provide a single LR that samples across all populations.

- **Stutter Variance (k^2) -** Parameter that describes how variable stutter peak heights are within the run

- **Total LR-** total LR for each population. Contributors within H1 (H_p) and H2 (H_d) are assumed to be unrelated individuals.

- **Unified LR –** LR that takes into account that the unknown contributors within H1 (Hp) and H2 (Hd) are made of up of both relatives and unrelated people

- **Weighting or Weight** – A probability that reflects how well a particular genotype combination explains the evidence profile. For example, if a proposed combination of genotypes is unlikely to lead to the observed evidence profile then that combination will be given a low weighting (close to zero).

Form Approved for use by:

DNA Technical Leader