



Le applicazioni dei marcatori molecolari nella ricerca in campo vegetale sono numerose e diversificate a seconda della natura genetica e della rilevanza biologica dei polimorfismi da questi evidenziati



Nell'ambito di una specie, gli individui si diversificano l'uno dall'altro per un numero più o meno elevato di caratteri (**alleli**) e questo mette nella condizione di poter rilevare polimorfismi (**mutazioni**) nelle regioni di DNA omologhe (**loci**)

I MM costituiscono strumenti eccezionali a disposizione della ricerca genetica poiché consentono di identificare con certezza una specifica sequenza nucleotidica e quindi analizzare polimorfismi di particolari geni o regioni cromosomiche

Attualmente le applicazioni più interessanti dei marcatori molecolari riguardano

Caratterizzazione e dissezione della variabilità genetica

Tipizzazione e identificazione varietale

Costruzione di mappe genetiche

Mappaggio genico e MAS

### Scelta della giusta metodologia

nella scelta della metodologia da applicare nei propri studi molti sono i punti da considerare:

#### Che tipo di informazione si vuole ottenere?

È importantissimo scegliere la giusta tecnica in modo da ottenere la giusta informazione. Per ottenere informazioni sulla genealogia di una popolazione o sulle relazioni filogenetiche, dovrebbero ottenersi delle sequenze o utilizzare metodi basati sulla restrizione del genoma

#### Discriminazione

a quale livello tassonomico la variazione genetica viene misurata  
(entro popolazione, fra specie, fra generi?)

il metodo che si sta scegliendo ci permette di avere informazioni dettagliate sul tipo di discriminazione che vogliamo ottenere?

### Scelta della giusta metodologia

#### Robustezza e riproducibilità

Alcune delle classi di marcatori più utilizzate (specialmente i RAPD) sono stati criticati per la loro mancanza di robustezza ed affidabilità.

Infatti molte riviste non prendono più in considerazione i lavori fatti con i RAPD

Recentemente questi problemi sono stati notati per molte classi di marcatori dimostrando che nessuna di esse è perfetta

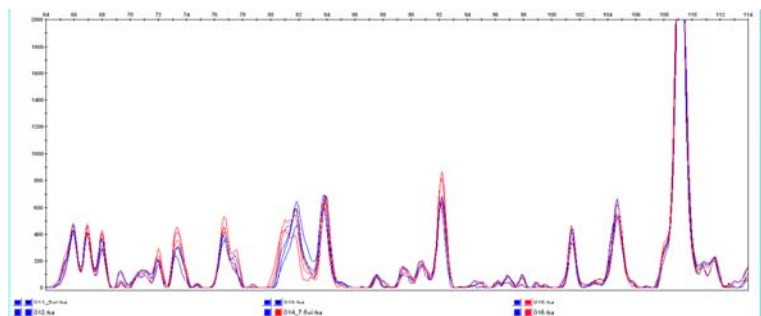
#### Reliability

Al fine di aumentare l'affidabilità del dato, molti autori propongono la ripetizione dei vari step dei protocolli (amplificazione ed elettroforesi) per 3 o 3 volte

Lo scoring è meno riproducibile del resto, per cui è importante avere due persone che fanno lo scoring in maniera separata

*Competitive priming* = problema che porta all'insorgenza di false bande

Differenze nel profilo di amplificazione AFLP tra radici e foglie di uno stesso genotipo ha permesso di evidenziare l'importanza della qualità del DNA da analizzare così come la dipendenza dal pattern di metilazione dei vari organi



Ma in genere questo non è un problema perché difficilmente queste tecniche vengono utilizzate per comparazioni tra laboratori

Per contro gli SSR sono considerati il metodo da scegliere per questo scopo

Gli SSR vengono considerati molto riproducibili

### Band homology - co-migrazione

I marcatori multi-locus sono stati utilizzati in moltissimi studi sulla similarità genetica

Assunto: bande che co-migrano rappresentano uno stesso tratto cromosomico

Ma non è sempre così....

Si ha "Size homoplasy" quando 2 o + bande della stessa lunghezza e ottenute con una stessa combinazione di primer, non hanno origine comune

Molto comune per i RAPD

osservato, però, anche con AFLP per frammenti piccoli

molto frequente per SSR a causa dell'alto tasso mutazionale e può divenire un problema quando gli SSR si usano per studi fra inter-specifici (es *Arabis* e *Arabidopsis*)

### Numero di loci

vengono richieste informazioni su pochi o molti loci?

RFLP, isoenzimi etc danno informazioni su pochi loci mentre RAPD, AFLP etc. danno una panoramica allargata a moltissimi loci.

### Costo

gli isoenzimi sono i più economici seguiti da RAPD e RFLP. Gli AFLP, SAMPL, S-SAP e i metodi basati sul sequenziamento sono i più costosi.

Se le sonde, o i primer specifici, non sono precedentemente disponibili, tecniche come SCAR e SSR risultano altamente costose perchè coinvolgono lo sviluppo di primer/sonde prima di poter essere applicate.

### Velocità

I metodi basati sulla PCR danno sicuramente risultati più rapidi laddove i primer sono già disponibili. I metodi basati sull'ibridazione sono invece più lenti.

### Impieghi

I marcatori molecolari possono essere di grande utilità quando utilizzati per

**verificare il modo di riproduzione** (sessuale vs. apomittico)

**verificare il sistema di unione** (autofecondazione vs. fecondazione incrociata),

**identificare varietà coltivate**

**per stimare le distanze genetiche**

sia per studi evolutivisti e tassonomici che caratterizzare germoplasma

## Analisi della variabilità (similarità e diversità) genetica e del flusso genico



Grazie all'introduzione di tecniche molecolari accurate e potenti è possibile raccogliere informazioni circa il grado di diversità o di similarità genetica di molte specie vegetali e la strutturazione della variabilità sia entro che tra popolazioni

I MM sono in grado di rilevare la variabilità di regioni di DNA omologhe in diversi individui nell'ambito di una popolazione o di più popolazioni della stessa specie o di specie diverse

Il modello teorico che consente di collegare le differenze riscontrate a livello molecolare tra taxa distinti con l'evoluzione delle specie è quello dell'orologio molecolare:

Assunta una velocità di mutazione costante, due sequenze di DNA (o due alleli) relative ad uno stesso locus di due individui differenti saranno tanto più diverse tra loro (in termini di numero di mutazioni accumulate) quanto maggiore è il tempo intercorso dalla loro separazione da una sequenza ancestrale appartenente a un loro progenitore comune

In sostanza, quanto maggiore è il numero di mutazioni a più loci portate da due individui, tanto maggiore è la diversità genetica tra loro



tanto maggiore è il numero di mutazioni accumulate dai due individui,  
tanto minore è la loro similarità genetica

### Modello mutazionale stepwise di Valdes

Prevede eventi mutazionali con uguale probabilità in una specifica sequenza nucleotidica, alleli di dimensioni simili risulterebbero simili anche in termini di entità degli eventi mutazionali subiti



due individui o due popolazioni devono considerarsi geneticamente distanti quando a più loci presentano alleli di dimensioni molto diverse (o con molte mutazioni)

In accordo con la teoria della neutralità di Kimura, però, la maggior parte delle mutazioni è silente (mutazioni neutrali) e dal punto di vista selettivo non ha effetti sulla fitness

I loci marcatori saggiati di solito sono selettivamente neutrali, cioè non hanno un significato particolare nell'adattamento degli individui all'ambiente

La caratterizzazione della variabilità genetica è realizzata mediante analisi statistiche dei dati molecolari, semplici da condurre e che forniscono risultati facili da interpretare



Questo prevede la stima dei coefficienti di similarità genetica ( $S_G$ ) o di dissimilarità genetica ( $D_G$ )

I dati relativi ai marcatori molecolari, raccolti nel campione di individui analizzati, vengono utilizzati per costruire le matrici di similarità o dissimilarità calcolando i coefficienti relativi in tutte le possibili combinazioni a coppia tra tutti gli individui di una popolazione e/o di più popolazioni

I coefficienti possono essere:

Jaccard (1908)

Dice (1945)

Nei e Li (1978)

Rohlf (1993)

Lo studio della struttura genetica delle popolazioni richiede, invece, analisi più complesse e articolate. Queste analisi prevedono, in primo luogo, il calcolo delle frequenze ( $p$  e  $q$ ) degli alleli marcatori a tutti i loci saggiati

Lo studio della struttura genetica delle popolazioni richiede, invece, analisi più complesse e articolate



Queste analisi prevedono il calcolo delle frequenze ( $p$  e  $q$ ) degli alleli marcatori a tutti i loci saggiati

Il calcolo è diverso in funzione della natura genetica del tipo di marcatori molecolari analizzati

**Marcatori codominanti:** per ciascun locus le frequenze alleliche possono essere determinate:

$$p_{a1} = (D + \frac{1}{2}H) \longrightarrow q_{a2} = 1 - p_{a1} = (R + \frac{1}{2}H)$$

dove  $D$  e  $R$  sono le frequenze dei genotipi omozigoti (per l'allele  $a_1$  e per l'allele  $a_2$ ),  $H$  è la frequenza del genotipo eterozigote  $a_1a_2$  che presenta entrambi gli alleli marcatori

Queste tre frequenze, con i marcatori codominanti, possono essere desunte direttamente dai risultati dell'esperimento di ibridazione o di amplificazione



ad un determinato locus genomico i genotipi omozigoti ( $a_1a_1$  e  $a_2a_2$ ) presentano una sola banda mentre i genotipi eterozigoti ( $a_1a_2$ ) presentano due bande corrispondenti agli alleli  $a_1$  e  $a_2$

**Marcatore dominante:** le frequenze alleliche possono essere derivate impiegando la seguente equazione:

$$p_{a1} = 1 - \sqrt{1 - F_b}$$

dove  $F_b$  è la **frequenza della banda relativa al locus** (che comprende situazione eterozigote e omozigote,  $a_1a_0$  e  $a_1a_1$ )

in una situazione di **equilibrio Hardy-Weinberg**

$1 - F_b$  rappresenta la **frequenza del genotipo omozigote ( $a_0a_0$ ) per l'assenza della banda**

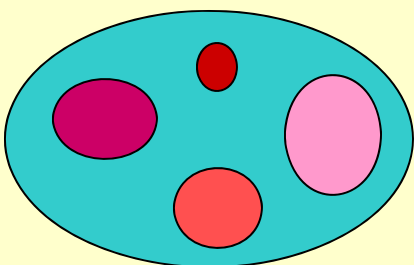
La radice quadrata del valore  $1 - F_b$  fornisce quindi la **frequenza dell'altro allele allo stesso locus ( $q_{a0}$ )**, quello che determina l'assenza della banda.



Wright (1965) ha elaborato un approccio per **valutare la struttura genetica delle popolazioni e la ripartizione della diversità tra sottopopolazioni**

sviluppando statistiche **valide per un locus con due alleli o per un locus con più alleli** considerando **come primo allele** quello più comune e **come secondo allele** tutti gli altri insieme

Le popolazioni naturali **possono essere considerate come un insieme di sottopopolazioni** distribuite in una determinata area geografica e tra loro parzialmente isolate dal punto di vista riproduttivo.



La **suddivisione di una popolazione provoca inbreeding** cioè unioni tra individui imparentati in misura maggiore rispetto a quanto voluto dal caso in una popolazione di dimensioni infinite

Ciò è dovuto al **campionamento non casuale dei gameti** che contribuiscono alla generazione successiva in sottopopolazioni di dimensioni ridotte

Tale situazione determina un eccesso di omozigosità nelle sottopopolazioni rispetto a quanto atteso in base alla legge dell'equilibrio di Hardy-Weinberg

il suo effetto sulla struttura genetica delle popolazioni può essere misurato in termini di diminuzione delle frequenze degli eterozigoti

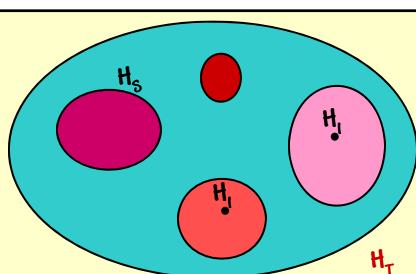
L'eterozigosità di una popolazione può essere studiata a tre diversi livelli di complessità

i) singoli individui (I)

ii) sottopopolazioni (S)

iii) popolazione totale (T)

Il grado di eterozigosità corrispondente a questi tre livelli è indicato come:  $H_I$ ,  $H_S$  e  $H_T$



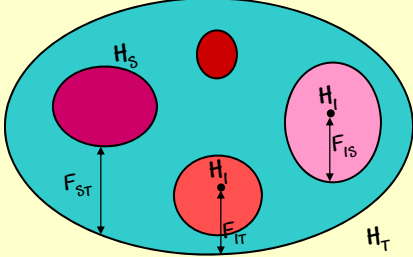
$H_I$  equivale all'eterozigosità media osservata negli individui delle sottopopolazioni

eterozigosità media a tutti i loci di un individuo e può essere interpretato come la probabilità che un individuo appartenente ad una sottopopolazione sia eterozigote per un qualsiasi gene

$H_S$  rappresenta il livello di eterozigosità attesa nella sottopopolazione

$H_T$  è l'eterozigosità attesa nella popolazione complessiva ottenuta riunendo tutte le sottopopolazioni

Le statistiche di Wright comprendono una serie di parametri  $F_{IS}$ ,  $F_{ST}$ ,  $F_{IT}$ , equivalenti a stime del coefficiente di *inbreeding*, che differiscono tra loro a seconda delle popolazioni di riferimento



FIS misura l'*inbreeding* degli individui rispetto alla sottopopolazione alla quale appartengono e calcolato come

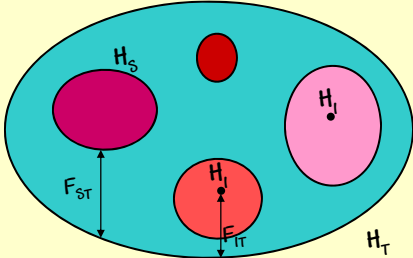
$$FIS = \frac{\overline{H_S} - H_I}{\overline{H_S}}$$

Questo parametro equivale alla **riduzione di eterozigosità media di un individuo a causa degli incroci non casuali all'interno della sottopopolazione**

Il secondo parametro, chiamato anche indice di fissazione e determinato come:

$$FST = \frac{\overline{H_T} - H_S}{\overline{H_T}}$$

misura gli **effetti della strutturazione della popolazione in sottopopolazioni** e **rappresenta la riduzione di eterozigosità di una sottopopolazione rispetto alla popolazione totale**



Nel 1973 Nei ha **esteso l'analisi a più loci** elaborando al contempo una **formulazione di coefficienti validi anche per loci multiallelici**

I primi due parametri, **HT e HS**, **equivalgono agli indici di diversità genetica totale della popolazione come complesso delle n sottopopolazioni campionate e di diversità genetica presente entro le singole sottopopolazioni per tutti i loci saggiati**

Dalla differenza tra questi due indici può essere ottenuta la **differenziazione genetica**:

$$DST = HT - \overline{H_S}$$

Il **grado di differenziazione (GST)** può essere derivato come **rapporto tra la differenziazione e la diversità genetica totale**

$$GST = \frac{DST}{HT}$$

Tale parametro, detto anche indice di fissazione, assume valore nullo quando le sottopopolazioni considerate sono in equilibrio ed hanno le stesse frequenze alleliche

L'indice di fissazione consente anche di determinare il flusso genico ( $Nm$ ) tra le sottopopolazioni in esame.

Il flusso genico può essere ricavato secondo la seguente formula:

$$Nm = \frac{c(1-GST)}{GST} = \frac{c(1-FST)}{FST}$$

dove  $c$  è pari a 0,5 per marcatori dominanti e a 0,25 per marcatori codominanti

Questo parametro assume valori inferiori ad 1 in assenza di flusso genico, mentre valori maggiori di 1 indicano presenza di flusso genico legato allo scambio di alleli marcatori tra le sottopopolazioni considerate.

Il flusso genico tra popolazioni di una specie coltivata generalmente può verificarsi:

attraverso la dispersione di polline tra campi confinanti, a cui fa seguito la fecondazione di ovocellule e quindi la produzione finale di seme vitale

attraverso lo scambio di seme tra agricoltori che riproducono in proprio la semente per la stagione successiva in diversi appezzamenti di terreno

### Caratterizzazione individuale (genotyping) e varietale (DNA fingerprinting)

Il fine ultimo del lavoro di miglioramento genetico è quello di **ottenere varietà superiori**, aventi le caratteristiche desiderate e rispondenti ad un ideotipo prefissato.

Il termine **varietà** indica un insieme di piante coltivate contraddistinte per caratteri morfologici, fisiologici, citologici, chimici, molecolari, ecc. che quando riprodotte, per via sessuale o asessuale, nei modi indicati dal costitutore, conservano i loro caratteri distintivi

*"D.U.S. testing" (Distinctiveness, Uniformity, Stability).*

La nuova varietà deve quindi possedere **caratteristiche peculiari che permettano di distinguerla da tutte le altre varietà della stessa specie** già iscritte al Registro Nazionale o al Catalogo Comune Europeo delle varietà.

Le **differenze** possono essere di **natura fisiologica** (ad esempio, epoca di fioritura/spigatura o di maturazione, e presenza/assenza di determinati composti, principi nutritivi e specifiche resistenze) oppure **morfologiche** (ad esempio, presenza/assenza di peli o di reste, e colore delle foglie o dei fiori).

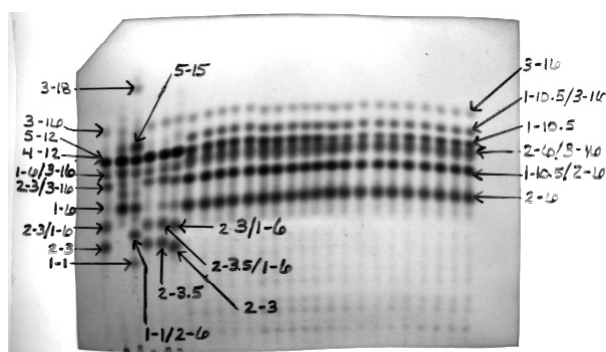


I requisiti di distinguibilità, uniformità e stabilità possono essere verificati con relativa facilità nelle varietà fondate su singole linee pure (varietà di specie autogame) o su singoli genotipi (ibridi semplici, varietà di specie a propagazione vegetativa).

Nelle specie allogame se le varietà sono in equilibrio Hardy-Weinberg (varietà locali, varietà costituite mediante selezione massale, selezione fenotipica o selezione ricorrente e varietà sintetiche), e nelle specie prevalentemente autogame per le varietà ottenute mediante selezione massale, i problemi che sorgono nell'accertamento dei requisiti D.U.S. sono molto più complessi.

Un carattere quantitativo può essere definito solo attraverso il calcolo della media e della deviazione standard (o dell'errore standard), parametri il cui valore può cambiare sensibilmente col variare delle condizioni ambientali. Ciò significa che la variabilità osservata è ampiamente soggetta ad oscillazioni da un ambiente ad un altro, così che una varietà mostra la variabilità minima nella sua zona di adattamento, mentre allontanandosi dal proprio ambiente la variabilità aumenta.

In alcuni casi, per caratterizzare le varietà di piante coltivate e quindi verificare la loro distinguibilità possono essere impiegati anche i marcatori biochimici, quali isoenzimi e proteine di riserva. Ad esempio, a tale scopo sono stati impiegati i sistemi isoenzimatici delle fosfo-glucosio-isomerasi (PGI) in loietto e delle esterasi (EXT) in pomodoro



Da più parti viene auspicato un maggiore uso dei marcatori biochimici e molecolari per la caratterizzazione varietale

Per le **autogame** gli isoenzimi e soprattutto le proteine di riserva possono essere sufficienti, unitamente ai caratteri morfologici, per individuare differenze tra varietà fondate sulla linea pura, ad eccezione forse delle specie che presentano moltissime varietà iscritte.

Il **problema principale** rimane l'identificazione di specifiche varietà e la distinzione tra varietà diverse dagli ibridi nelle **allogame**.

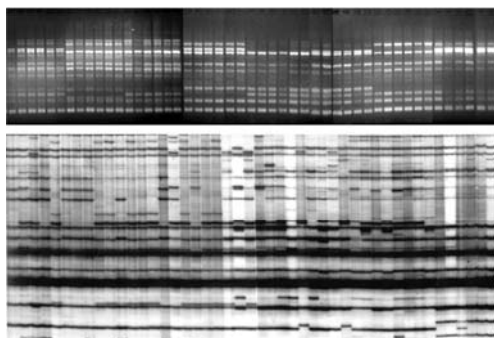
In questo caso è auspicabile l'adozione delle nuove metodologie basate sulla rilevazione di marcatori molecolari, sia RFLP che PCR-derivati (SSR, AFLP).

Il **problema principale** che la Commissione Sementi del Ministero è chiamata a risolvere è, comunque, la distinguibilità, che può essere considerato l'elemento centrale del sistema D.U.S.

Le prospettive migliori per la caratterizzazione e l'identificazione varietale sembrano comunque essere offerte dagli strumenti biotecnologici.

In questi termini non possono essere usate le medie dei caratteri quantitativi e nemmeno le frequenze dei marcatori molecolari (anche se con analisi multivariate ed analisi discriminanti è possibile ottenere informazioni utili) poiché non sarebbero in grado di fornire una risposta certa ed univoca, avente valore non soltanto scientifico ma anche giuridico.

In questo contesto potrebbero essere seguiti due approcci principali. Il **primo**, prevede il "tagging" dei materiali di base, ad esempio i cloni di una sintetica o le inbred di un ibrido, mediante transgeni non senso o polimorfismi molecolari rari.



esempio di caratterizzazione varietale basati sui profili elettroforetici generati dai prodotti di amplificazione.

Il **secondo** è, invece, basato sul "fingerprinting" al fine di **individuare marcatori polimorfici discriminanti tra varietà in termini di presenza/assenza**. In questo ultimo caso i marcatori più affidabili sono probabilmente gli RFLP ma quelli più informativi sembrano essere gli AFLP ed i più riproducibili gli SSR.

Il DNA fingerprinting è stato introdotto la prima volta nell'analisi del genoma vegetale nel 1988



Ryskov et al. distinguono 2 genotipi di orzo utilizzando una sonda M13 su DNA digerito con HaeIII

Ryskov et al. (1988). FEBS letter 233:388-392

Le moderne **tecniche molecolari di analisi del genoma** possono anche essere applicate per la protezione dei Diritti del Costitutore (*Plant Breeder's Rights*).

Tuttavia, bisogna tenere presente che la normativa vigente nella UE in materia di brevetti afferma chiaramente che **le varietà vegetali costituite secondo i metodi tradizionali di selezione sono escluse da tale sistema di protezione giuridica**

Il sistema di caratterizzazione molecolare **finalizzato all'identificazione varietale** deve essere potenziato per tutelare i costitutori, visto anche che per il principio del libero accesso alle risorse genetiche, possono essere ottenute nuove varietà a partire da varietà già iscritte, note come **varietà essenzialmente derivate** (*essentially derived varieties*).

In questo caso, l'accertamento della derivazione e della minima distanza genetica tra le varietà richiederà necessariamente l'adozione di metodi di analisi a livello molecolare e di appropriati coefficienti di similarità o diversità genetica.

## Casi di studio



La lenticchia di Castelluccio di Norcia

Specie prevalentemente autogama con quota d'incrocio inferiore all'1%

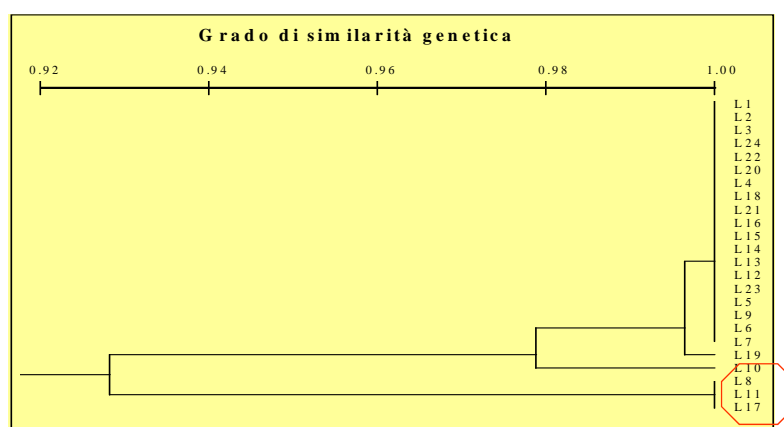
La lenticchia è una specie che presenta un'elevata variabilità genetica: ciò è dimostrato dalla presenza di numerose varietà locali sul territorio italiano.

La lenticchia di Castelluccio di Norcia è la più pregiata tra le varietà locali di *L. culinaris* coltivate in Italia



è l'unica che ha ottenuto il marchio IGP

## Caratterizzazione morfo-fisiologica e molecolare



Frode commerciale