1 Introduction

1.1 Aims of this project

Escherichia coli typically colonizes the avian gastrointestinal tract and other mucosal surfaces. While most strains are commensal, certain strains designated Avian Pathogenic *Escherichia coli* (APEC) have the ability to cause severe disease. Predominant serotypes of APEC are O1:K1, O2:K1, and O78:K80 (Blanco *et al.*, 1998; Dozois *et al.*, 1992; Glantz *et al.*, 1962). APECs most likely enter and colonize the avian respiratory tract by inhalation of faecal dust, causing localized infections which lead to airsacculitis and pneumonia. In certain cases, they spread into various internal organs and typically cause pericarditis, perihepatitis, peritonitis, salpingitis, and other extraintestinal diseases. In its acute form the disease is characterized by septicemia commonly resulting in sudden death (Barnes and Gross, 1997).

Several bacterial factors have been associated with the virulence of APEC, including adhesins, toxins, iron acquisitions systems, Colicin V plasmid, serum resistance proteins, capsules as well as the lipopolysaccharide complex. However, various factors have only been identified by infection models like one-day-chicken models or application of APEC directly into the airsac, which circumvent the natural route of infection (Dho-Moulin and Fairbrother, 1999; Ewers *et al.*, 2003). Therefore, the mechanisms underlying natural pathogenesis are still not fully understood and only certain steps of the infection process can be explained by these already known virulence factors.

In recent years, applications of genome-wide analyses have led to a better understanding of the molecular mechanisms of pathogenicity. New molecular approaches have also aided in the identification of genes involved in pathogenesis, including In vivo expression technology (IVET), Selective capture of transcribed sequences (SCOTS), Differential fluorescence induction (DFI), or Signature-tagged transposon mutagenesis (STM) (Graham and Clark-Curtiss, 1999; Hensel *et al.*, 1995; Mahan *et al.*, 1993; Valdivia and Falkow, 1996). Among these technologies, STM is the most successfully and widely used. It allows for

large scale of screen, and selected mutants are directly associated with virulence attenuation.

The objective of this study is to identify and clone virulence factor genes critical for septicemia at large scale, through which we can indirectly confirm the actual role of the already known virulence factors during pathogenesis and simultaneously identify novel genes of virulence factors. This approach is expected to give a more deeply insight into the molecular mechanism of the pathogenicity of APEC infections and to hopefully lead to new molecular approaches to prevent diseases caused by APEC. Therefore, the STM strategy is utilized in this study to accomplish the following four aims:

Aim 1: establish a new and clinically relevant infection model;
Aim 2: adapt STM to this newly established infection model;
Aim 3: identify genes disrupted by transposon insertion, that attenuate virulence in vivo;
Aim 4: genetically characterize selected mutants.

1.2 Current knowledge of the molecular basis of APEC pathogenicity

Several potential virulence factors have been identified in APEC, mainly from positive correlation between phenotypic characteristics and virulence for chickens. Analysis of the involvement of these factors in virulence using experimental models of infection just began during the past few years. This involvement includes the adherence ability of bacteria, the resistance of bacteria to immunological defence mechanisms, the multiplication ability of bacteria in the host, and the ability to produce cytopathic effects. Nevertheless, all the actual roles of the known virulence factors are not fully elucidated and certain steps of the infection process have not been related to previously identified factors.

1.2.1 Type 1 fimbriae

Type 1 fimbriae are prevalent not only in APEC, but also in UPEC (uropathogenic *E. coli*) and EHEC (enterohemorrhagic *E. coli*) as well as in most commensal isolates. These organelles consist of a major subunit FimA and several ancillary subunits such as FimF, FimG and the adhesin FimH. All subunits are encoded by the *fim* gene cluster, harbouring a total of nine genes, seven of which are present in a single operon whose expression is controlled by an invertible element containing the promoter (Orndorff, 1994). The assembled pili consist of a 7 nm thick helical rod (made up of the major subunit) coupled to a 3 nm wide tip fibrillum structure containing the adhesin FimH and two adaptor proteins, FimF and FimG (Jones *et al.*, 1995; Russell and Orndorff, 1992). Several variants of type 1 pili have been described on APEC and they seem to be related to the serotype of the strains. Four variable regions have been identified in *fim*A of an APEC isolate and at least two regions are specific for isolates of the O2-group (Dho-Moulin *et al.*, 1990; Marc and Dho-Moulin, 1996; Suwanichkul and Panigrahy, 1986). For the ligand subunit FimH, studies were focused on enterohemorrhagic *E. coli* and uropathogenic *E. coli*. A number of natural and engineered allelic variants have been identified over the last several years. They differ by as little as one single amino acid and confer distinct adhesive phenotypes to type 1 fimbriae (Sokurenko *et al.*, 1994; Sokurenko *et al.*, 1995; Sokurenko *et al.*, 1997; Sokurenko *et al.*, 1998).

The significance of type 1 pili in pathogenesis of APEC has been demonstrated by previous studies. The presence of type 1 fimbriae is more frequent in pathogenic than in non-pathogenic avian *E. coli* strains, even though these fimbriae are common among *E. coli* (Dozois et al., 1992). It was shown that a virulent and fimbriated strain was cleared less efficiently from the trachea of turkeys than an avirulent and less fimbriated strain (Arp and Jensen, 1980). Virulent *E. coli* strains were better colonizers of the chicken trachea than avirulent strains, and this ability to colonize was mediated by type 1 fimbriae (Dho and Lafont, 1982). Adhesion of type 1 fimbriae to chicken epithelial cells of the pharynx was demonstrated both in vivo and in vitro (Dho and Lafont, 1984; Naveh et al., 1984). Gyimah and Panigrahy (1988), blocked specific adherence of APEC strains to chicken tracheal sections with specific anti-type 1 fimbriae serum and by D-mannose, which is the cellular receptor of the adhesin of type 1 fimbriae. Type 1 fimbriae may also play a role in the interaction of APEC strains with the immune system by blocking phagocytosis and increasing resistance to bactericidal effects of serum, although this role is controversial (Orndorff *et al.*, 1996).

However, the actual role of type 1 fimbriae during infection is unclear. By using a mutant deleted for the entire *fim* operon, type 1 fimbriae were shown not to be essential in the colonization of the trachea and airsacs, but could play a role in the colonization of the lungs instead (Marc *et al.*, 1998). In addition, it has been demonstrated that the FimH adhesin is not required for bacterial colonization of the trachea in germ-free chickens. On the contrary, the absence of FimH seems to favour colonization in the trachea of chickens. This adhesin does also not appear to be necessary for manifestation of bacterial pathogenicity in these chickens. Clearly, further studies are required to elucidate the authentic role of type 1 pili during pathogenesis of APEC.

1.2.2 P fimbriae

Similar to type 1 pili, P fimbriae are also composite organelles consisting of a short, flexible tip fibrillum attached to the distal end of a thicker rod structure. The pilus rod is composed of repeating main subunits of PapA anchored to the outer membrane possibly by the PapH subunit (Baga *et al.*, 1987). PapG, which is thought to be the adhesin, in association

with three other subunits, PapE, PapF, and PapK, constitutes the tip fibrillum (Jacob-Dubuisson *et al.*, 1993; Kuehn *et al.*, 1992). At least 11 serotypes of P fimbriae, F7-F16, identified based on antigenic differences in the major subunits and three distinct PapG variants, designated GI, GII, and GIII, have been identified. These variants recognize different isoreceptors of a series of glycolipids. The minimal glycolipid receptor for the PapG adhesin is termed globotriasylceramide (GbO3) which consists of a digalactoside (Galα1-4Gal) core linked by a β-glucose (Glc) residue to a ceramide group. Alteration of the Galα1-4Gal core by the addition of a single or two N-acetyl-galactosamines (GalNAc) generates GbO4 or GbO5. Adhesin variants GI, GII, and GIII recognize GbO3, GbbO4, and GbbO5, respectively (Feria *et al.*, 2001).

P fimbrial adhesins were initially found on UPEC strains. Studies have indicated that P fimbriae mediate bacterial adherence to uroepithelial cells and are significant virulence factors associated with pyelonephritis. Certain APEC strains also express P fimbrial adhesins. However, in general, these adhesins have been observed in a low proportion of the isolates studied. In studies performed by Dozois *et al.* (1992) and Dozois *et al.* (1995), expression of P fimbrial adhesins was detected by mannose-resistant haemagglutination (MRHA) and immunofluorescence only in O1 isolates from septicemic turkeys and in an O18 isolate from septicemic chicken in a total of 112 *E. coli* isolates, whereas others demonstrated that 52% of the O2 isolates from septicemic chickens expressed P fimbrial adhesins as detected by MRHA (Achtman *et al.*, 1986). Moreover, it was reported that 78% of 203 isolates from chickens with septicemia, or 96% (when only isolates of serotypes O1:K1, O2:K1, O35 and O78:k80 were considered) expressed P fimbrial adhesins of type F11, as detected by ELISA (van den Bosch *et al.*, 1993). Interestingly, 44% of the isolates from septicemic chickens and 31% isolates from septicemic turkeys harboured *pap*-related DNA sequences (Dozois *et al.*, 1992), a much higher proportion than that of expression of P fimbial adhesins. Similarly, 22.7% and 30.0% isolates from internal organs of poultry having died from colisepticemia were *papC* positive as detected by PCR and hybridization (Janßen *et al.*, 2001; Ewers, 2004).

Thus, the role of P fimbrial adhesins in the pathogenicity of APEC has also not yet been fully elucidated. These adhesins do not seem to confer bacterial adherence to chicken tra-

chea or pharyngeal cells (van den Bosch *et al.*, 1993; Vidotto *et al.*, 1997) and this fimbrial adhesin was not expressed in bacteria colonizing the trachea of inoculated chickens. However, specific F11-antibodies were stimulated by inoculation of an F11 P-fimbriated APEC strain into chicken via trachea or airsack (Pourbakhsh *et al.*, 1997b). This provided the evidence that P fimbriae are expressed in vivo. P fimbriae were also shown to be expressed in bacteria colonizing the airsack, lungs, and internal organs. These results suggested that P fimbriae mediated adhesion is not important during the initial colonization of upper respiratory tract, but rather in the later stage of infection.

1.2.3 Iron uptake systems

Iron is essential for bacterial growth which requires a concentration of about 10^{-6} mol. l^{-1}. However, only low amounts of free iron (10^{-18} mol. l^{-1}) are present in body fluids and most of this iron is bound to physiological carriers, mainly transferrins. For this reason, bacteria developed two strategies to sequestrate this essential element. Firstly, bacteria express high affinity receptors (Hanson *et al.*, 1992). With these receptors, bacteria can utilize transferrin, lactoferrin, and even heme degraded from hemoglobin as iron source. Secondly, bacteria have developed high affinity systems of iron acquisition, such as synthesis of siderophores which are able to compete with iron-chelators of the host (Ratledge and Dover, 2000).

One of the most efficient siderophore present in several species of enteric organisms is aerobactin, a hydroxamate low molecular-weight compound able to chelate iron with rather high affinity. The aerobactin system is encoded by 5 genes, organized in one operon and four of these genes, *iuc*A, *iuc*B, *iuc*C, and *iuc*D, are responsible for the synthesis of aerobactin. The fifth gene, *iut*A, encodes a membrane receptor which binds the aerobactin-iron-complex (Neilands, 1992). The aerobactin operon is usually carried by large *col*V plasmids, which are of at least 80 kb in size (Waters and Crosa, 1991).

The significant role of aerobactin in the pathogenesis of APEC has been suggested by several previous studies. A positive correlation was demonstrated between expression of the aerobactin system and lethality of *E. coli* isolates for chickens (Lafont *et al.*, 1987). Several studies have confirmed that most APEC strains (73-98%) possess and express aerobactin,

regardless of their serotype, whereas non-pathogenic strains produce aerobactin far less frequently (Dozois *et al.*, 1992; Emery *et al.*, 1992; Linggood *et al.*, 1987). However, the role of aerobactin in avian colispeticemia has not yet been fully clarified. The aerobactin operon from an APEC strain was transferred to various avirulent *E. coli* strains, but the recombinant strains were not lethal for one-day-old chickens, and moreover the antibody against aerobactin could not protect chickens challenged with virulent strains, expressing aerobactin (Goes *et al.*, 1993).

Evaluation of these experiments is further complicated by the fact, that not all APEC strains possess the aerobactin system, but can anyway grow in iron-limited environments. Thus, some other iron uptake systems seem to be able to take over the task of iron sequestration. Besides typical enterobactin and aerobactin, a third siderophore was identified, yersiniabactin, which is encoded by the so-called 'high pathogenicity island' (HPI) (Karch *et al.*, 1999). In previous studies, almost all isolates of O-group O78 and O2 possessed *irp*2 and *fyu*A, two critical genes found on this HPI, and about 70% of APEC isolates, regardless of their serotype, carried both the HPI and aerobactin in parallel (Gophna *et al.*, 2001; Janßen *et al.*, 2001). However, the significance of yersiniabactin in the pathogenesis of APEC remains to be shown. More recently, another two systems Sit (Salmonella iron transport) and Chu (*E. coli* haem-utilization) have been identified in APEC strains. However, no direct evidence was shown that they are involved in the virulence of APEC by these authors (Schouler *et al.*, 2004; Stocki *et al.*, 2002).

1.2.4 Tsh protein

The Tsh protein was first identified in the O78 APEC strain χ7122. This large protein conferred a temperature-sensitive hemagglutination, preferentially expressed at low temperatures (26-30°C), and repressed at 42°C (Provence and Curtiss, 1994). This serine protease is a member of the autotransporter group of proteins. Its sequence exhibits similarity to the immunoglobulin A proteases of *Neisseria gonorrhoeae* and *Haemophilus influenzae*, as well as to several other proteins of pathogenic bacteria. Purified Tsh is capable of adhering to red blood cells, hemoglobin, and the extracellular matrix proteins fibronectin as well as collagen IV. It also was demonstrated that Tsh exerts proteolytic activity against casein,

and serine 259 is essential for the protease functions. However, this reside is not required for adherence to substrates, and its replacement by an alanine does not abolish binding activity (Kostakioti and Stathopoulos, 2004; Otto *et al.*, 1998). A sequence diversity between Tsh and yet another autotransporter, Hbp (hemoglobin protease) is only two amino acids, the latter being able to degrade hemoglobin of human and release haem from haemoglobin (Otto *et al.*, 1998).

Tsh is synthesized as a 140-kDa precursor protein, whose processing results in a 106-kDa passenger domain (Tsh_s) and a 33-kDa ß-domain ($Tsh_ß$). Tsh_s-protein is the secreted part, while $Tsh_β$-protein is located in the outer membrane (Stathopoulos *et al.*, 1999). In general, like most other genes coding for autotransporter proteins, *tsh* is located on large plasmids, often on a *col*V plasmid.

Maurer *et al.* (1998) demonstrated that Tsh was associated with APEC, but not with *E. coli* strains isolated from the feces of healthy chickens. The frequencies with which *tsh* was detected in several studies present relatively large differences. In a previous study investigating 300 avian *E. coli* isolates originating from France and Quebec, *tsh* was found in 90.6% of the pathogenic, and only in 3.4% of non-pathogenic isolates (Dho-Mouli *et al.*, 1997). Janßen *et al.* (2001) also demonstrated that 85.3% of strains isolated from colisepticemia were *tsh* positive. However, in studies by Dozois *et al.* (2000), Delicato *et al.* (2002) and Ewers *et al.* (2004), *tsh* was detected in only 49.7%, 39.5% and 53.3% of APEC strains. These differences may be due to the similarity of Tsh to a newly identified virulence factor: vacuolating toxin (Parreira and Gyles, 2003). Recently, experiments demonstrated that Tsh is not required for later systemic development of avian septicemia, but does only contribute to the development of airsack lesions in the early stages of infection (Dozois *et al.*, 2000).

1.2.5 K1 capsule

Some bacterial species produce extracellular polysaccharide or other polymers such as capsule. The chemical composition of capsule varies according to the bacterial species under investigation. In *E. coli*, at least 80 capsule polysaccharides are known. Capsule

types K1, K2, K3, K5, K12 and K13 are frequently associated with pathogenic *E. coli* strains (Orskov and Orskov, 1992).

Capsular polysaccharides have been shown to be essential virulence determinants of *E. coli* causing extra-intestinal infections in mammals. For example, the K1 capsule is associated with 80% of the isolates that cause *E. coli* human neonatal meningitis, and *E. coli* strains expressing K1 comprise the majority of isolates from neonates with septicemia and acute pediatric pyelonephritis (Robbins *et al.*, 1974). The K1 antigen is also frequently associated with APEC of pathogenic O-groups O1, O2, as well as non-typeable APEC strains (Gross, 1994b).

By analogy with results obtained with *E. coli* of human origin, the K1 antigen is thought to be involved in the resistance of APEC to the immune defenses of the birds. Pourbakhsh *et al.* (1997a) showed that three APEC strains possessing the K1 antigen were more resistant to the bactericidal effect of serum than other APEC strains expressing various different K antigens. A K1 mutant was also demonstrated to become sensitive to the bactericidal effect of complement and completely lost pathogenicity in different experimental infection models tested. More recently, investigations with a spontaneous K1-mutant demonstrated that the production of capsule in APEC strain MT78 endows the bacteria with decreased colonization potential in infected chickens and furthermore is responsible for increased resistance to the bactericidal effects of chicken serum and phagocytosis by interaction with complement system (Mellata *et al.*, 2003a; Mellata *et al.*, 2003b).

1.2.6 Lipopolysaccharide (LPS)

LPS is one of the major components of the bacterial outer membrane. This amphipathic molecule is composed of lipid A, a core oligosaccharide, and a polysaccharide chain termed O-specific antigen. General features of lipid A and the core oligosaccharide are highly conserved, reflecting constraints imposed by their essential roles in outer membrane integrity as well as its ancient phylogeny. The polysaccharide chain provides the most striking example of structural diversity and so far over 170 distinct antigens have been reported for *E. coli* (Heinrichs *et al.*, 1998). This may be due to the selective pressure caused by interaction between the O-antigen chain and the host immune system.

One potential outcome of gram-negative infections is septic shock, a syndrome manifested by hypotension, coagulopathy, and organ failure. Septic shock results from the liberation of LPS from the bacterial cell surface, a phenomenon that is naturally from growth and proliferation of bacteria. The so called endotoxin does not damage tissue directly, but up-regulates host production of cytokines and inflammatory mediators, thus leading to tissue and vascular damage and the symptoms of sepsis (Proctor *et al.*, 1995).
Clones corresponding to genes involved in the synthesis of LPS have recently been identified by selective capture of transcribed sequence (SCOTS) analysis (Dozois *et al.*, 2003). The O78-antigen of the APEC strain χ7122 used in this study was previously shown to be required for virulence by increasing the bacterial resistance against serum (Mellata *et al.*, 2003a; Mellata *et al.*, 2003b). The O-polysaccharide also contributes to the virulence of many pathogenic bacteria by inhibition of phagocytosis and the alternative complement fixation pathway (Joiner, 1988).

1.2.7 Resistance to the bactericidal effects of serum

A strong association between serum resistance and isolates from septicemic chickens has been demonstrated by previous studies. This resistance is mediated by bacterial surface structures such as outer membrane proteins, ColV production, Iss (increased serum survival), capsule and LPS. The outer membrane protein A (OmpA) and TraT (Transfer T) play a special role in serum resistance. An $OmpA^+$ isolate showed 10-fold higher survival rates in a 10-day-old chicken embryoned egg model than did the respective $OmpA^-$ strain. Moreover, this $OmpA^-$ mutant was demonstrated to be sensitive to bactericidal effects of serum (Weiser and Gotschlich, 1991). TraT, an outer membrane protein encoded by resistance plasmids, was also reported to enhance resistance of *E. coli* to phagocytosis (Aguero *et al.*, 1984). However, on the contrary, Mellata *et al.*, (2003a) demonstrated that TraT, as well as Iss, may play a limited role in the defence against killing by serum.

1.2.8 Toxins and cytotoxins

It was recently reported that APEC produce a cytotoxin that is similar to the *Helicobacter* (*H.*) *pylori* vacuolating cytotoxin A (VacA) (Salvadori *et al.*, 2001). The cytotoxic substances in supernatants of both APEC and *H. pylori* cause similar morphological changes in target cells and are similar with respect to heat lability and susceptibility to proteolytic enzymes. DNA sequencing has subsequently revealed that *vat* is located in a novel described pathogenicity island of APEC strain Ec222 (O78) (Parreira and Gyles, 2003). The predicted 148.3-kDa protein shares 75% protein identity to the hemagglutinin Tsh and possesses three domains that are typical for serine protease autotransporters of *Enterobacteriaceae*: an N-terminal signal sequence of 55 amino acids, a 111.8 kDa passenger domain containing a modified serine protease site (ATSGSG), and a C-terminal outer membrane translocator of 30.5 kDa. A Δ*vat* isogenic mutant of Ec222 showed no virulence in a respiratory and cellulitis infection model of disease in broiler chickens.
There are also a few reports that APEC produces verotoxins, heat-labile toxin as well as shigatoxin. But the respective proteins or genes have not yet been characterized and their presumable roles in pathogenicity thus could not be confirmed so far (Babai *et al.*, 1997; Blanco *et al.*, 1997).

In general, while having identified several virulence-associated factors during the last decade, for mainly three areas of reasons it has not been possible to elucidate their role during the pathogenesis unambiguously: on the one hand several different strains harboring different genomic contents have been used. This fact prevents comparison of findings between single experiment outcomes. In addition, these different strains have been used in various different animal infection models. And thirdly, the role of some virulence-associated genes has so far only been shown in vitro or via epidemiologically investigating a limited population of APEC.

1.2.9 Molecular techniques applied for identification of virulence-associated genes in APEC

In addition to the characterization of putative virulence genes and the study of their role in pathogenesis, several novel genomic methods have been used to study the pathogenesis of APEC infections. Brown and Curtiss (1996) performed a genomic subtraction between APEC strain χ7122 (serotype O78:K80) and a K-12 *E. coli*. These experiments enabled these authors to identify and locate 12 unique regions on the chromosome of this particular APEC strain. By using a similar approach, Coulange (1999) isolated 17 DNA fragments specific for a pathogenic strain. More recently, Suppression subtractive hybridization (SSH) has also been used successfully to identify genes present in the genome of two APEC strains, but absent in *E. coli* K-12 MG1655 (Dozois *et al.*, 2003; Stocki *et al.*, 2002). Dozois *et al.* (2003) applied SCOTS to identify conserved genes in APEC strain χ7122 that are expressed in infected chicken tissues. Pathogen-specific DNA or cDNA, encoding putative adhesins, lipopolysaccharide core synthesis, iron-responsive, plasmid- and phage-encoded genes and genes of unknown function were successfully enriched and isolated by these authors. These findings underline the strength of genome wide screening approaches for the identification of new virulence factors, which will greatly improve our knowledge of the pathogenic mechanisms of APEC infections.

1.3 Strategies for the identification of novel virulence factors

One hundred years after Koch, Stanley Falkow (Foulongne *et al.*, 2004) proposed a new, molecular version of Koch's postulates to define specific virulence genes, the so-called Falkow´s molecular postulates: (a) a certain gene confers a certain phenotype to the studied bacteria, (b) inactivation of this gene abolishes the phenotype, (c) re-introduction of this particular gene restores the mutant to wild type. Although this strategy, based upon mutagenesis and the use of experimental models, allows the identification of several genes, it is not comprehensive but only leads to the characterization of this specific gene´s role in the respective model. Thus, fulfilment of Falkow´s postulates resembles the definite proof of gene involvement in disease.

However, other methods are needed to complete de novo identification of virulence factors such as differential expression, either at the level of transcription (transcriptome) or at the level of protein expression (proteome). All these techniques are now supported by data from complete genome sequencing projects. The pool of information obtained from these approaches allows the definition of the 'virulome', which is the assembly of factors a pathogen requires for virulence. Understanding the virulome will path the way to the development of new strategies for vaccination or the development of new generations of antimicrobial agents. Therefore, novel technologies such as Signature-tagged transposon mutagenesis (STM), In vivo expression technology (IVET), Suppressive subtractive hybridization (SSH), In vivo induced antigen technology (IVIAT), Selective capture of transcribed sequences (SCOTS) and bacterial proteomic analysis will be briefly reviewed.

1.3.1 Signature-tagged transposon mutagenesis (STM)

Several different approaches have been used to exploit transposon mutagenesis for the isolation of bacterial virulence genes. The “traditional” “one-mutant, one-animal” method is generally labor intensive, expensive, and otherwise unwieldy. These limitations were circumvented by developing a transposon mutagenesis system termed Signature-tagged mutagenesis (STM) in which each transposon was tagged with a known specific DNA sequence or tag (Hensel *et al.*, 1995). This tag is prepared as a pool, each tag consisting of

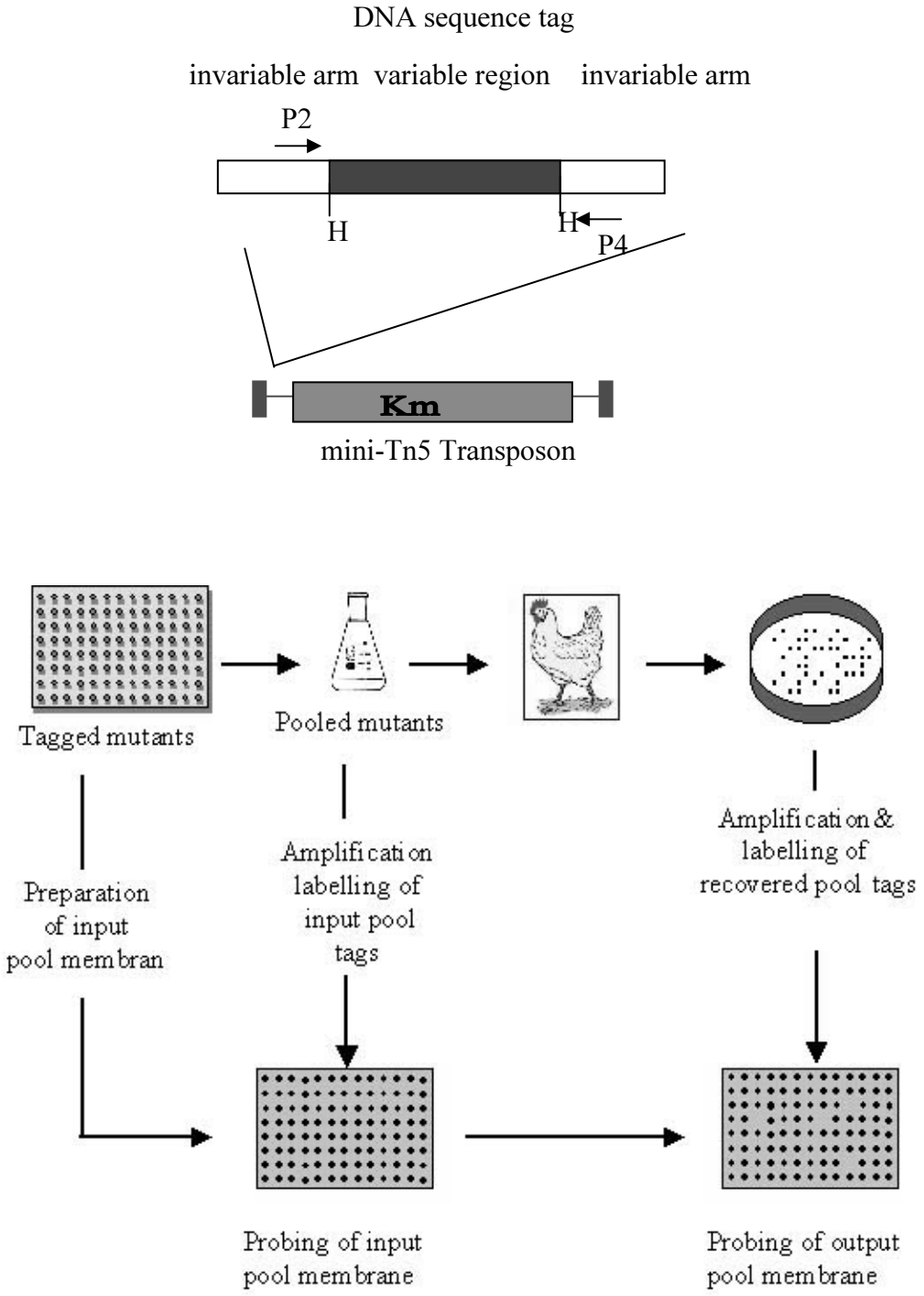

Figure 1: STM used for simultaneous identification of *Salmonella* Typhimurium virulence genes based on transposon mutagenesis and negative selection. Abbreviations and symbols: P, primers; H, *Hind*III; Km, kanamycin resistance gene.

a variable central region and regions flanked by arms of invariant sequences by oligonucleotide synthesis and PCR. Each tag comprises an individual sequence of 40 bp ($[NK]_{20}$; N=A, C, G or T; K=G or T) and prevents the occurrence of sites for the restriction enzymes *Hind*III, *Kpn*I, *Pst*I, as well as *Sal*I, but is sufficiently variable to ensure that the same sequence should occur only once in 2×10^{17} molecules. These double stranded tags are then ligated into the mini-Tn5*km2* transposon and can be transferred from donor *E. coli* to recipient bacteria by conjugation. Thus, bacteria are mutated by transposon mutagenesis. Each mutant has a unique molecular signature, which can be specifically detected by using PCR with primers specific for the outer common arms. Animals are infected with pools of mutant clones, and any reduction in virulence, namely failure to reproduce in the host, indicates that the insertion has disrupted a gene critical for the growth and survival of the pathogen in vivo. Those clones present in the input pool, but absent in the recovered pool, are identified by a colony blot technique. The original STM method was subsequently modified to avoid the tedious pre-screening process and allows use of nonradioactive hybridization (Fuller *et al.*, 2000).

When screenings in more than one infection model are performed by the "traditional" STM methodology, a high number of hybridizations are needed. Struve et al. (2003) developed a novel multi-screening Signature-tagged mutagenesis (MS-STM) assay to ease simultaneous screenings of a mutant bank in more than one infection model (Struve *et al.*, 2003). In the MS-STM assay, the PCR-amplified tags present in the inoculum and in the recovered pools are spotted onto membranes and separate hybridizations are performed with each specific tag as a probe. Therefore, the number of hybridizations necessary to screen a mutant library in more than one infection model is significantly reduced, since pools of tags recovered from more than one infection model can be analysed simultaneously. Moreover, fewer numbers of PCR amplifications are required when the MS-STM assay is applied, because two rounds of PCR are no longer needed to obtain labelled tags. In addition, the presence of invariant arms does not result in higher background levels of hybridization in the MS-STM assay, in contrast to the invariant arms, which must be removed before spotting the amplified tags onto the membranes in the "traditional" STM strategy.

In theory, STM was designed to find general applicability to both gram-positive and -negative, animal and plant bacterial pathogens. The recent modifications and rapid development of STM technology now make general applicability feasible. The major advantage of STM is that it directly identifies genes involved in virulence at a large scale rather than by any other genome-wide technology. Moreover it does not depend on selection parameters and can be used to identify genes that are expressed transiently or at low levels. However, it should also be noted that STM might not identify mutations causing small or even moderate reductions in survival depending on the sensitivity of the PCR/hybridization protocol for detecting such changes in tag population (Chiang *et al.*, 1999). For example, mutations in genes of some toxins might not affect bacterial survival in vivo, although they are critical for causing disease.

1.3.2 In vivo expression technology (IVET)

Responding to the change of circumstance, many pathogenic organisms modulate their patterns of gene expression by up or down-regulating those that are specifically required for survival in the host. At least some of these in vivo induced (ivi) genes are supposed to play a critical role in pathogenesis and several promoter-trap strategies were used to identify genetic loci whose expression is induced in host environments. Such techniques are generally termed "in vivo expression technology'' (IVET) methods.

It is known that purine auxotrophs of *Salmonella enterica* serotype Typhimurium are rapidly eliminated from mice unless they are complemented. The original version of IVET took advantage of this fact to develop a promoter-trap (Mahan *et al.*, 1993). Random fragments of the *S.* Typhimurium chromosome were ligated with a promoterless *purA-lacZY* synthetic operon to form transcriptional fusions. This library was introduced into Δ*purA* strain by conjugation to generate merodiploids via homologous recombination at the chromosomal fragment. The mutant bank was injected into mice and positively selected because only strains expressing the gene fusion in vivo can survive. The recovered bacteria from mice were screened for *lacZY* expression on lactose-MacConkey indicator medium to avoid the subsequent study of strains containing constitutively active fusion. Similar IVET strategies, incorporating *purA* or other selectable complementing genes, have been applied

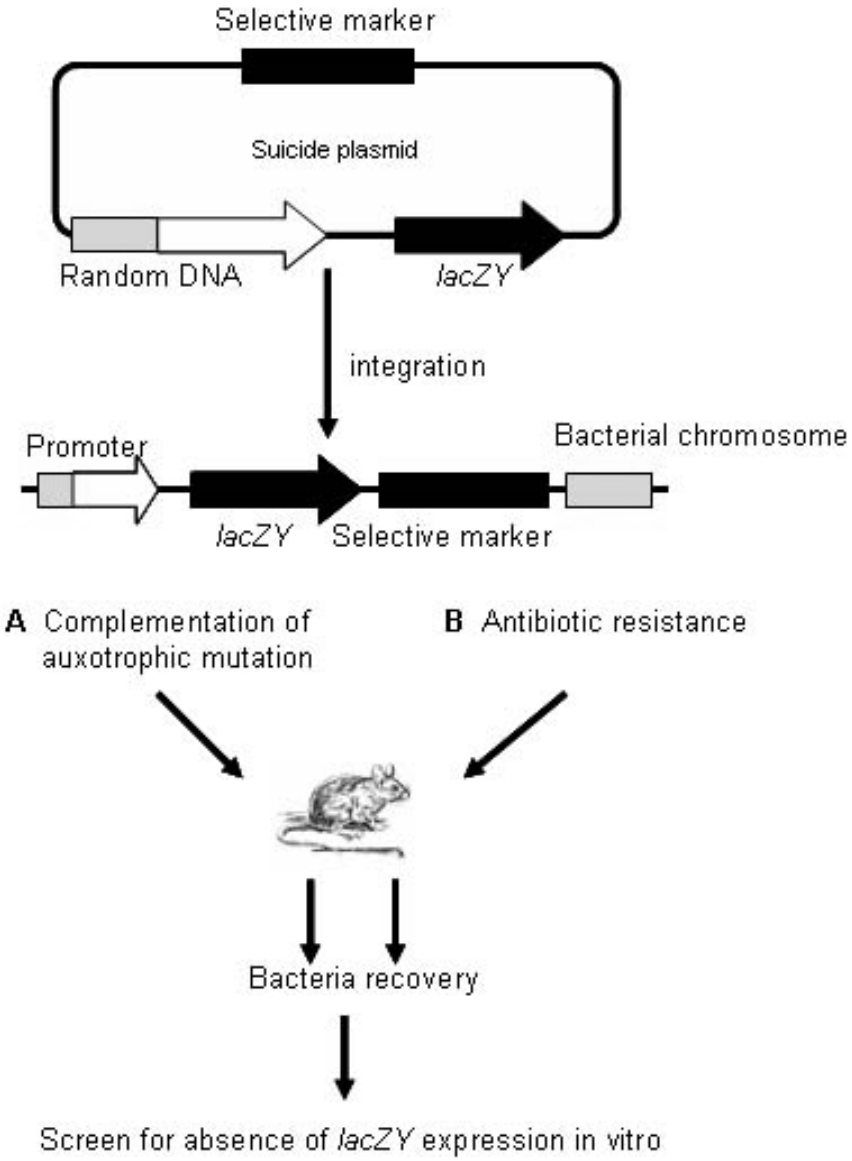

Figure 2: IVET system developed to positively select for genes induced specifically in the host (the first and second version of IVET). Abbreviations: *lac*ZY, promoterless *E. coli lac* operon genes (Buer and Balling, 2003).

in a large number of gram-negative and gram-positive pathogens, as well as in a fungal pathogen. The major limitations of this version of IVET selection are that genes transiently expressed or low level expressed are difficult or even impossible to be identified.

Since the attenuating and complementable auxotrophy is not available for all microbial systems, a variation of IVET on the basic principle was developed. In this second version of IVET, antibiotic resistance genes serve as selectable reporter (Mahan *et al.*, 1995). Using this antibiotic based method, it should be possible to carry out selection for ivi genes in any tissue in which the antibiotic concentration can be made sufficiently high to select against strains not expressing the resistance gene. In addition, ivi promoters with different levels of activity may be isolated by adjustment of the antibiotic dosage; and ivi genes that are expressed at particular time or location during infection may be identified by variation of the timing of antibiotic administration. However, a disadvantage of this approach is that the appropriate antibiotic must be administered to host animals and must penetrate to the site of infection.

The third version of IVET is a recombination-based IVET (RIVET) (Camilli *et al.*, 1994). In this case, fusions are made to a promoterless resolvase gene such as *tnpR* from Tnγδ. TnpR catalyses irreversible recombination between specific DNA sequences, termed as *res* sites. By constructing a system in which resolvase activity results in permanent excision from the chromosome of an antibiotic resistance gene flanked by *res* sites, this event marks the bacterium by endowing it with an inheritable antibiotic-sensitive phenotype. Resolved strains are then screened after recovery of the bacteria from infected tissues. Because only a small pulse of resolvase expression is needed to mediate excision, RIVET is exquisitely sensitive to low or transient expression of the ivi gene during infection. Another advantage of this method is that no selective pressure is imposed upon the bacteria during infection and thus guarantees to proceed on a natural infection course. RIVET has been modified to study the regulation of ivi genes, which is difficult for standard methods. An interesting alternative use of RIVET is the monitoring of induction of ivi genes as a function of time and location in the host.

A promising new method for identifying genes induced during infection is differential fluorescence induction (DFI) (Valdivia and Falkow, 1996). DFI is yet another variant of

IVET using a single *gfp* gene as a dual reporter, although it is listed independently from IVET by some authors. In the original application, random fragments of chromosomal *S.* Typhimurium DNA were fused with a promoterless *gfp* gene, and the resultant library was introduced into *S.* Typhimurium. FACS was used to isolate mutants containing *gfp* fusion that are up-regulated in vivo but down-regulated in vitro. One of the advantages of DFI is automation for initial screening compared to manual screening; the other is that DFI allows the study of upregulation as opposed to on-off types of gene induction.
To sum up, the IVET technology has become a utility for virulence gene discovery in many pathogens and a stimulus for the creation of new tools to investigate pathogenicity.

1.3.3 Selective capture of transcribed sequences (SCOTS)

Selective capture of transcribed sequences (SCOTS) is a technique that can be used to evaluate gene expression by microoganisms in a diversity of environments, in vivo and in vitro (Daigle *et al.*, 2001; Graham and Clark-Curtiss, 1999). Total RNA is isolated from infected tissue and then treated with RNase-free DNase I. The pooled RNA samples from different individual animals are then converted to first-strand cDNAs by random priming with unique terminal sequences, the resultant cDNAs are made into double-stranded with Klenow fragment. Subsequently, cDNA is amplified by PCR using the defined primers for each set of cDNA. In order to selectively capture the transcribed sequences, denatured, biotinylated, and sonicated demonic DNA fragments of interesting pathogens are mixed and hybridized with sonicated ribosomal RNA. After preblocking by ribosomal RNA, denatured amplified cDNA from infected tissue are added and hybridized. Bacterial cDNA that hybridized to biotinylated genomic DNA are retained by binding hybrids to streptavidin-coated M280 magnetic beads. The captured cDNA is then eluted, precipitated, and amplified by PCR.
SCOTS does not require well-characterized genetic tools to manipulate the organism of interest, or extensive knowledge of genome of the organism. It is applicable to any microorganism from which nucleic acids can be isolated. In addition, SCOTS uses wild type strains without any genetic modification and can provide insights about host-pathogen interations, as with other methods used for identification of in vivo gene expression.

1.3.4 Genome analysis and mapping by in vitro transposition (GAMBIT)

The GAMBIT approach is a versatile system for creation and phenotypic analysis of pools of transposon mutants that combine in vitro transposon mutagenesis and genetic footprinting. It was originally developed for functional genomic studies of two naturally transformable bacterial pathogens, *Haemophilus influenzae* and *Streptococcus pneumoniae*, but is applicable to any haploid organism that can be efficiently transformed with linear DNA (Akerley *et al.*, 1998). A specific region of the chromosome is amplified by long-range PCR, and its product is subjected to in vitro transposon mutagenesis. The resultant pool

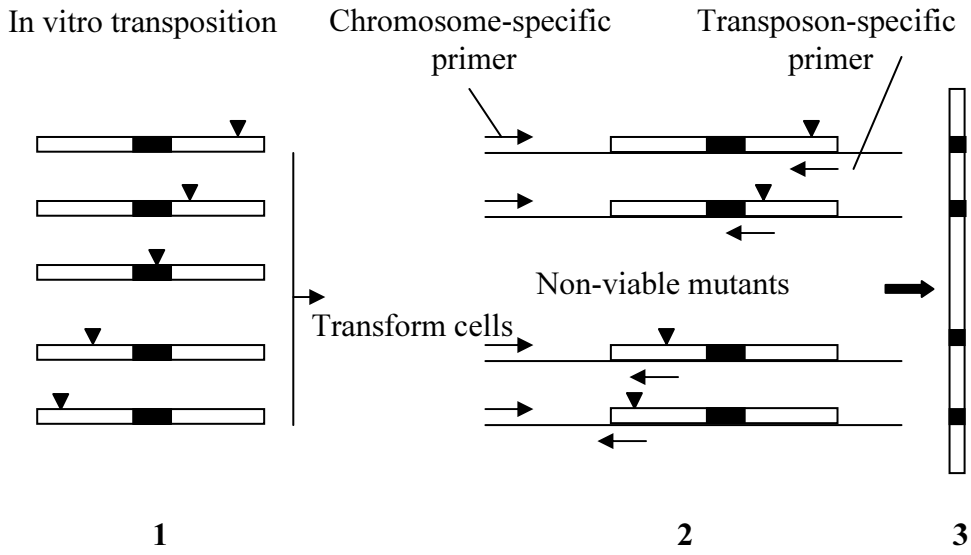

Figure 3: The GAMBIT procedure adapted from Akerley and Lampe (2002) (Copyright 1998, National Academy of Science, U.S.A.). Random *mariner* insertions (triangles) in a linear DNA target are generated by in vitro transposition **(1).** Mutant pools are created by transformation into naturally competent bacteria and selection for an antibiotic resistance gene carried by the transposon followed by PCR of the complex pool **(2),** and gel electrophoresis to map insertion sites **(3).** Regions that do not sustain insertions in vivo are putatively essential for growth or viability.

of mutagenized DNA is then transformed into bacteria, which are then grown under selective conditions (e.g. on defined medium or in an animal host). PCR is subsequently performed on the postselection pool using a transposon-specific primer and a primer to a known location on the chromosome. Analysis of the PCR products allows determination of which genes in that region of chromosome are required for survival under those selective conditions.

Nowadays genome sequences of many pathogens have been revealed. A large number of regions of heterogeneity between bacterial strains including putative pathogenicity islands (PAIs) have been identified. A rapid means of mutagenizing these loci (short of designing deletion constructs on a gene-by-gene basis) is desirable for integrating the large databases of candidates produced by genome-scale approaches with the next step in functional characterization. Mutagenizing only the region of interest means that a more cumbersome large-scale screen is not needed. Besides regions of chromosome and pathogenicity island, large plasmids of several virulent pathogens are especially hit by this method. In addition, GAMBIT can identify some essential genes that are critical in vitro (Chiang *et al.*, 1999). They are main targets for antimicrobial strategies. STM is unable to assess the role of these genes in pathogenesis, because transposon insertions in these loci would lead to be lethal. IVET would equally not identify these genes unless they are expressed at extremely low levels in vitro.

1.3.5 DNA microassays

The sequence analysis of microbial genomes opens up new avenues to the understanding of bacterial pathogenesis. The genome sequence provides thousands of genes including both confirmed and putative. For example, even in the extensively studied *Escherichia coli*, functions of more than one third of the genes are unknown. High-throughput methods for assessment of function are clearly required if this wealth of primary sequence information should be used. Transcription control plays a key role in virulence regulation, thus genome-wide transcription profiling seems particularly appropriate for the study of this process.

DNA microarray–based approaches have gained rapid acceptance in a variety of fields for studying the roles of genes in pathogenesis of infectious disease. Designing a whole-genome DNA microarray for fully sequenced microbes is conceptually straightforward. Most of open reading frames (ORFs) can be quickly and accurately predicted by several sensitive microbial gene-finding programs. DNA fragments representing each of the ORFs can be obtained by PCR amplification with ORF-specific primers. Homology-searching algorithms should be utilized to choose regions of genes that will not cross-hybridize with other regions of the genome. After a simple purification step, PCR fragments can be arrayed by a robotic arrayer. The utility of microarray is not restricted to fully sequenced microbes. As applied for the eukaryotic pathogen, *Plasmodium falciparum* (Hayward *et al.*, 2000), a DNA microarray of 3.648 random genomic clones was used to identify more than 50 genes that were expressed differently between the trophozoite and gametocyte stages. Several powerful tools such as subtractive hybridization and suppression subtractive hybridization can be used to obtain the arraying DNA libraries. The major limitation of this approach is that the identity of any element of interest must be determined after the experiment.

A major challenge to all pathogens is adapting to changing environmental conditions. This adaptation involves regulating gene expression in response to different environmental signals. Expression of genes that are required in a particular situation will be up-regulated, whereas unnecessary functions in the same situation will be down-regulated. The same principle can also be applied to virulence genes, which are also subject to regulatory mechanisms that ensure expression in the appropriate host environment. However, measuring a pathogen's gene expression in microenvironments specific to the pathogen may be technically challenging (de Saizieu *et al.*, 1998). Because of the relatively lower number of pathogens present in the infected tissue, most RNA isolated from infected tissues is host specific RNA. And even in scarce amount of the pathogens RNA, most of it is ribosomal RNA. Until more sensitive detection protocols are developed, examining global gene expression will be more practical in environmental conditions that mimic aspects of the host environment, such as elevated temperature, iron limitation, and changes in pH and in cell culture models. Even after measurement of bacterial gene expression from infected hosts

became feasible, the ex-vivo datasets will be helpful to better understanding of the pathways of virulence factors regulation.

1.3.6 In vivo inducible antigen technology (IVIAT)

In vivo induced antigen technology (IVIAT) is a novel technology that can quickly and easily identify in vivo induced genes during infections (Hang *et al.*, 2003). This technology is expected to facilitate the discovery of new targets for vaccines, antimicrobials and diagnostic strategies in a wide range of microbial pathogens. A general overview of IVIAT is presented in Fig. 4. An inducible expression library of the pathogen has to be constructed, and sera from animals which have experienced disease caused by the pathogen of interest should be collected. The pooled serum is absorbed with whole cells and cellular extracts prepared from the pathogen grown in vitro. This absorbed serum is used to probe the inducible expression library described above by standard western blotting. Positive clones are identified by their position on the master plate, and each positive clone is purified at least two additional times and confirmed as reactive with adsorbed sera. Finally plasmids from individual reactive clones are purified, and the DNA inserted in the vector can be sequenced in both directions by using pET30-specific primers.

Pioneering technologies such as IVET, STM and SCOTS were designed to identify genes specifically expressed in vivo. Although significantly powerful, their major drawback is that they depend on the use of animal models of infection. However, animal models are not available for many pathogens or it might not closely approximate the real infection condition. The major conceptual and technological breakthrough is that IVIAT does not rely on animal infection models, as it utilizes serum from animals suffering from natural infections with the respective pathogen of interest (Handfield *et al.*, 2000). The use of pooled sera also allows to study the widest possible array of antigens produced during different stages of infection and when appropriate, from animals infected via different routes. Additional important advantages compared with other current strategies are that it is rapid, simple, and uses only routine immunological and genomic cloning methods to render a genome-wide screening in a timely fashion. It does not require prior knowledge of basic genetics of the

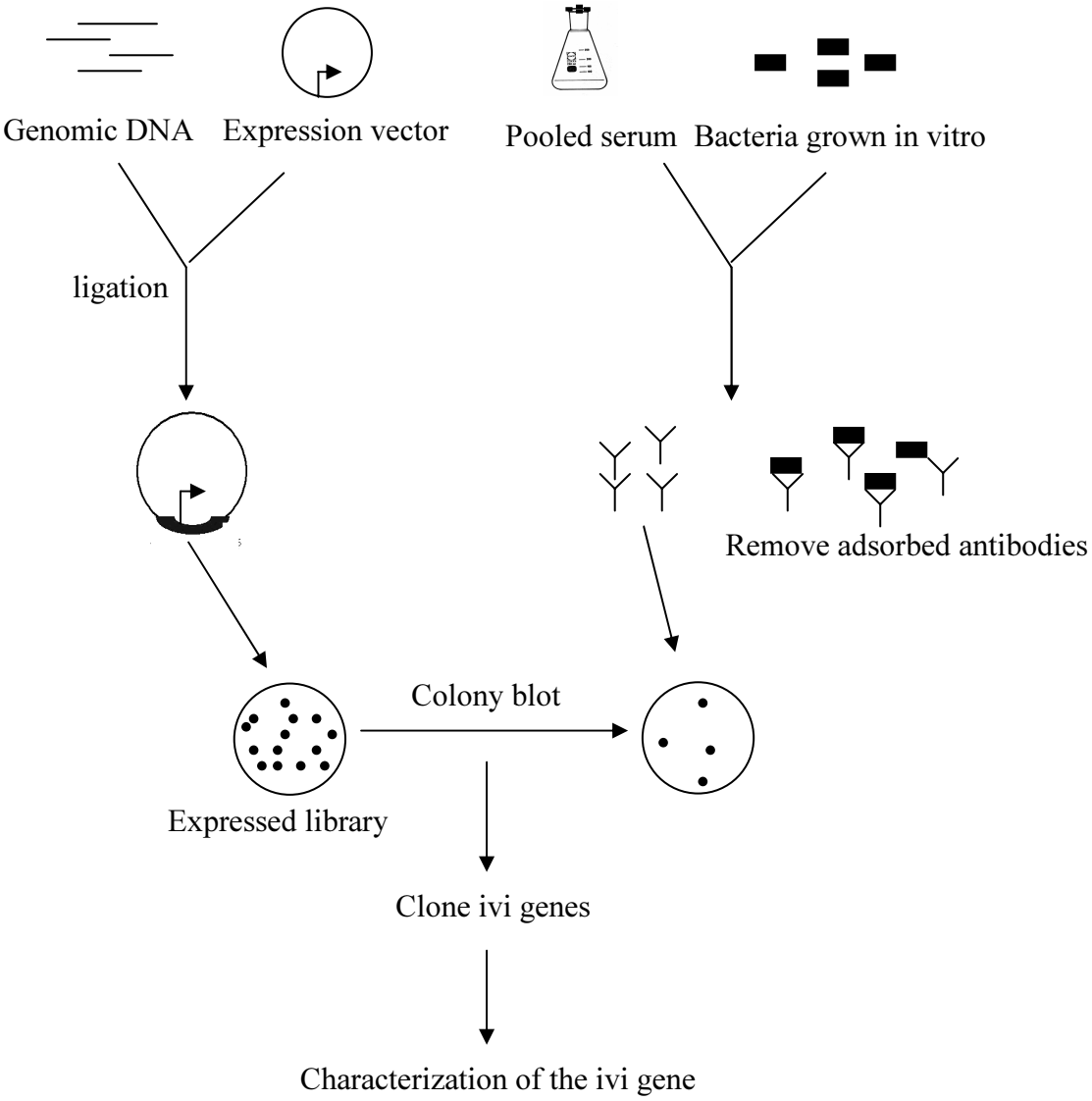

Figure 4: In vivo induced antigen technology (IVIAT) (Handfield *et al.*, 2000). Sera from animals that have experienced an infection caused by the pathogen under study are pooled and exhaustively absorbed with cells of the pathogen grown in vitro. An expression library of the pathogen's DNA is generated in a suitable host and clones are probed with adsorbed serum. Reactive clones, which are producing antigens that are expressed during a natural infection but not during in vitro cultivation, are purified and their DNA sequenced. Finally the ivi genes are characterized.

organism of interest. This is particularly useful in the study of new and emerging pathogens.

1.3.7 Proteome analysis

Monitoring gene expression at the mRNA level has many advantages: i) it has been easily scalable to cover near-to-total prokaryotic genomes; ii) it is readily automated by robots thus reducing sample handling and error. However, it is proteins not RNA to exercise biofunctions. Biological influences including the stability, half-life, post-transcriptional and degradative modification of proteins combine to suggest that mRNA abundance and response to environmental stimuli may not completely represent the expression of a corresponding protein in vivo. The global analysis of protein expression from a genome under a given set of conditions is best summarised by the term 'proteomics'. Studies on proteomics are essential for understanding pathogenesis, functional pathways, phenotype and elucidating biological functions of previous unknown proteins.

Two-dimensional gel electrophoresis (2DGE) remains a powerful method for proteomic study because of its high resolution, reproducibility, and sensitivity. In this approach, proteins are separated according to isoelectric point by electrofocussing in the first dimension and by molecular mass with sodium dodecyl sulfate electrophoresis in the second dimension, variations are also possible. Since these two parameters are unrelated, it is possible to obtain an almost uniform distribution of protein spots across a gel and in two dimensions.

The classical two-dimensional gel electrophoresis using isoelectric electrofocussing (IEF) in the first dimension and SDS-PAGE in the second has four core limitations: i) it is unable to solubilise and separate highly hydrophobic proteins prior to IEF; ii) there are no suitable buffers to reproducibly separate highly basic proteins; iii) the acrylamide pore size of both the first dimension IPG strip and second dimension slab gel excludes high mass proteins and protein complexes; and iv) it depends on the sensitivity of gel staining (Cordwell *et al.*, 2001). However, several strategies have been developed to overcome these limitations.

Recently, proteome analysis combining immunoblotting has led to development of immunoproteome technology. This new approach will facilitate the identification of new can-

didates for vaccines, as well as diagnostic and therapeutic tools (Jungblut and Bumann, 2002).

There are still some other new technologies reported, including inducible antisense RNA technology, tansposon footprinting, subtractive and surpression subtractive hybridization. Each of these approaches has advantages and disadvantages. All of these methods should contribute to an understanding of relevant physiological attributes displayed by bacterial pathogens in vivo. No single approach can provide all of the information necessary for understanding microbial pathogenesis. Each of these approaches, sometimes in combination with one or more other approaches, will provide much more comprehensive results in some cases but not in others.